Molecular Simulation Studies of Chromonic Mesogens in Aqueous Solution

THIND, ROMNIK

Molecular Simulation Studies of Chromonic Mesogens in Aqueous Solution

Romnik Thind

A thesis presented for the degree of
Doctor of Philosophy

Chemistry Department
Durham University
2017

Abstract

This thesis focuses on understanding the aggregation behaviour of lyotropic chromonic liquid crystals in aqueous solution. Molecular simulation methods are used to provide structural and thermodynamic information on the self-assembly of chromonic mesogens, and the results used to re-interpret data from previous experimental studies of these systems.

Extensive atomistic level molecular dynamic simulations have been performed on three chromonic dyes in solution: 5,5'-dimethyoxy-bis-(3,3'-di-sulphopropyl)-thiacyanine triethylammonium salt (Dye A), 5,5'-dichloro-bis-(3,3'-di-sulphopropyl)-thiacyanine triethylammonium salt (Dye B), and Bordeaux dye. The results are compared to key experimental data, such as X-ray scattering and cross-sectional areas and aggregation free energies. Previously suggested chromonic aggregation models, such as a double-width column and a brickwork layer structure, have been discounted based on the simulation results. Instead the simulations of dyes A and B demonstrate an anti-parallel stacking arrangement providing columns in which solubilizing sulphonate groups lie on alternate sides of the column as the column is traversed. In addition, a new type of chromonic "smectic" layered phase is predicted for these molecules. For dye A, a novel chiral column structure is seen within isolated columns in isotropic solution. For Bordeaux dye a stable "single-molecule" column is seen, along with a number of meta-stable structures including a double-width column in which two single-molecule columns are linked via a salt bridge.

In an attempt to simulate larger time and length scales, two *bottom-up* coarse-graining techniques, iterative Boltzmann inversion (IBI) and force matching (FM), were tested on simple hexane/water systems and then applied to Dye A. Both of these methods proved to be largely unsuccessful in reproducing the same aggregation behaviour seen in the atomistic simulation work.

Declaration

The material contained within this thesis has not previously been submitted for a degree at Durham University or any other university. The research within this thesis has been conducted by the author unless indicated otherwise.

Romnik Thind

2017

Acknowledgements

I would like to thank my supervisor, Prof. Mark Wilson, for his patience, support and encouragement during my studies. It has been a privilege to have worked with a supervisor who creates such a fun and dynamic working environment.

I would like to thank everyone from room 200 and 200X, where I feel I have learned as much about cake and tea as I have about science and research. A special mention must be given to Dr Martin Walker, whose constant support in research and thesis writing has been of immeasurable help.

I am grateful for the computer facilities provided by the Durham University High Performance Computing service and the support from the Engineering and Physical Sciences Research Council (EPSRC).

Finally, and most importantly, I would like to give a special thanks to my family and friends. The love and care from my uncle, brother and sister, in addition to the support and encouragement from all of my friends, has been a true blessing during my time in Durham.

Contents

List of Figures

List of Tables

Chapter 1

Introduction to liquid crystals

1.1 Liquid crystals

From the ancient Greek philosopher Empedocles, who claimed that all the structures in the world were composed of four elements (fire, air, water and earth), to the modern day plethora of divisions and terms; the definition and physical description of what matter is has changed and adapted throughout history. The diverse world of condensed matter increasingly requires knowledge of both the basic building blocks and also how they interact, to refine our understanding of the physical world around us.

For a long period of time, the states of matter believed to exist were limited to the most common states observed in the natural world: solid, liquid and gas. Over the years, the emergence of new and exciting phases has shown that states of matter are far more diverse and exotic than previously imagined. A multitude of structures are possible for specific systems when transitioning between a solid and a liquid phase. These mesophases result from exhibiting properties from both the crystalline solid and amorphous liquid phase.

Crystals have long and short range positional order between molecules, with the particles located in precise, regular, repeating patterns. A liquid phase has short range positional order, but, in sharp contrast to the crystalline phase, usually exhibits very rapid and disordered molecular motion, resulting in only short range positional order within the system. The liquid crystal phase shares key structural

characteristics and properties from both of these phases. The mesophase exhibits anisotropic behaviour, whereby the molecules show long range orientational order, denoted as a nematic phase. Under certain conditions, the mesophase can additionally have long range positional order between molecules, denoted as a smectic phase. A simple way of conceptualizing this type of structure is to picture an "ordered fluid".[7,8]

1.1.1 History and discovery

The discovery and history of liquid crystals is of remarkable interest. It is only in recent decades that liquid crystals have been heavily used in applications, revolutionizing the display industry in particular. It is perhaps then surprising that they were discovered over a hundred years ago. In the year 1888, at the Institute of Plant Physiology of the Germany University of Prague, an Austrian scientist Friedrich Reinitzer noted irregular behaviour when characterising a cholesteryl acetate and cholesteryl benzoate.[9,10] Observing what he termed a 'double melting', leading him to seek the help of German physicist, Otto Lehmann, to probe the optical properties of these molecules. Lehmann ascertained that the cholesteryl benzoate flowed like a conventional liquid, but refracted polarised light like a crystalline solid. This discovery marked the start of research into the field of liquid crystals.[11–13]

The history of liquid crystal research can be divided into three time periods of historical importance. The discovery in March 1888 led to increasing work in this field, with French crystallographer Georges Friedel using the difference in molecular organisation to categorise liquid crystals into three types of ordering: nematic, smectic and cholesteric. The time period of 1925-1950 showed very little progress, caused by not only the outbreak of both world wars, but also scientists failing to see any commercial uses for over 200 compounds documented as exhibiting liquid crystal behaviour. From the years 1958 to the present day, an international resurgence in liquid crystal research has been observed, causing an explosion in understanding the complexities of these phases. This was accompanied by a staggering increase in the number of liquid crystalline properties being exploited commercially. From an early application of using color-changing cholesteric compounds as temperature

sensors, to the development of Liquid Crystal Displays (LCD).[7,14] LCDs are perhaps the most well-known techonological application to date, having unrivaled utilization in an unfathomable amount of devices, from computer monitors and televisions, to watches and phones, to name just a handful.[7,14]

1.1.2 Types of liquid crystal

Liquid crystals can be divided into two main types: thermotropic and lyotropic. The mesogen, a term used for the individual molecule that forms a mesophase (in this case a liquid crystal phase), is typically an organic molecule.

Phase transitions between mesophases, characterised by the level of orientational and translational order within the system, can be heavily influenced by the environment they find themselves subject to. Transitions caused by a change in temperature only are termed a thermotropic phase, whereas a system that has phase transitions based on the alteration of both temperature and concentration (due to the presence of a solvent), are termed a lyotropic phase.[15]

1.1.3 Thermotropic phases

Thermotropic phases occur where phase transitions are dictated by a change in temperature of the system. The very specific molecular geometry of the mesogen drives the formation of a highly ordered system. The most common shaped mesogens are calamitic (rod-shaped) or discotic (disc-shaped), shown in Fig 1.1, discussed separately now.

For calamitic mesogens, the nematic phase (N) is characterised by having long range orientational order (mesogens are pointing on average in the same direction), but liquid-like positional order (shells of molecules around each mesogen). The average direction by which the molecules are oriented is called the director, $\mathbf{n}$, which, for calamitic molecules, aligns parallel with the long axis of the mesogen. The smectic phase (M) is characterised by having long range orientational order similar to the nematic phase, with additional 1-dimensional positional order. The molecules organize into layers. Comparing the two most common smectic phases, in a smectic-

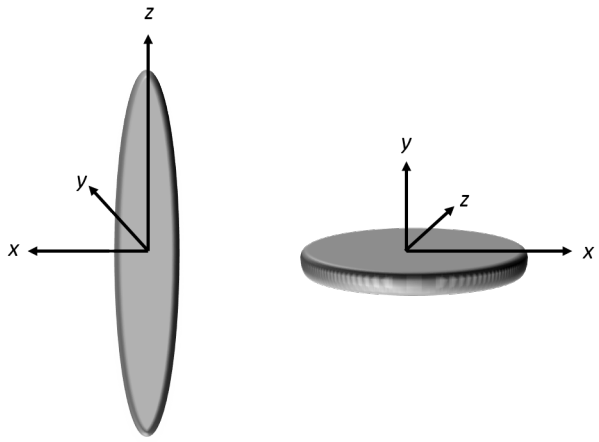

Figure 1.1: Schematic of calamitic and discotic mesogens for thermotropic phases.

A the director is parallel to the layer normal, while in the smectic-C phase, the director is tilted with respect to the layer normal.

For discotic mesogens, the discotic nematic phase is characterised by long range orientational order, but no long range positional order, with the director, **n**, lying parallel to the molecular short axis.[16] The discotic columnar phase is characterised by the discotic mesogens stacking to form columns (akin to stacks of pound coins), with these columns then orientating to form a two-dimensional lattice structure.[8]

The range of common thermotropic phases formed by calamitic and discotic mesogens can be seen in fig 1.2.

1.1.4 Lyotropic phases

Lyotropic liquid crystals, where the phase transitions are dictated by a change in both temperature and concentration, can be divided into two types based on the general molecular structure of the mesogen and aggregation properties: conventional amphiphilic and chromonic.[17]

Amphiphilc molecules

Amphiphilic molecules, more commonly known as surfactant molecules, are not only widely prevalent in nature (for example, providing pivotal structural support within a biological membrane in the form of a bilayer), but have been heavily used for over 1000 years in everyday applications, the most common usage being emulsifiers

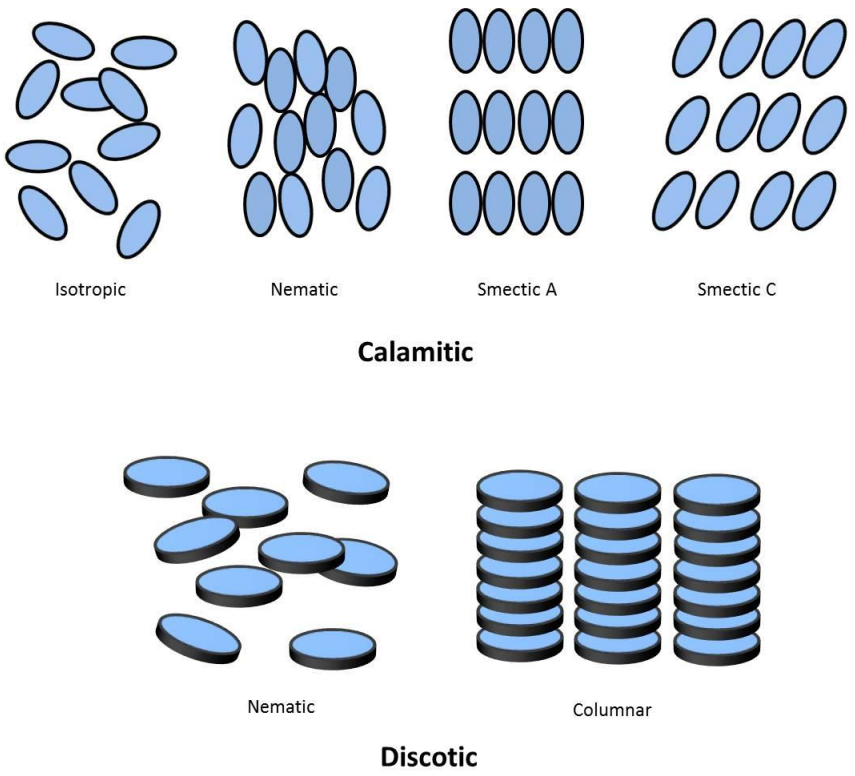

Figure 1.2: Schematic showing the common phases formed from calamitic and discotic mesogens.

in cleaning.[18] Surfactants have a general molecular structure characterised by a polar head group region, attached to a large non-polar hydrocarbon chain (commonly referred to as a hydrocarbon tail), as shown in fig 1.3. Solvation of these molecules, typically in water, leads to a unique result, due to the hydrophilic and hydrophobic regions of the molecule interacting in opposing ways in the solvent environment. While the polar head region thrives in a polar solvent environment, the non-polar hydrocarbon tail experiences unfavourable interactions, with resulting mesophase structures being the result of the ability of the amphiphilic molecules to form structures that minimise the interaction of the hydrocarbon tail with the solvent. The unfavourable reduction in entropy, caused by the aggregation of surfactant molecules, is counterbalanced by an increased overall entropy by the liberation of the solvation shells around the hydrophobic tail. At low concentrations, the mesogens aggregate to form spherical micelles, whereby the aggregated molecules position the polar head groups on the surface of a sphere, encapsulating the hydrocarbon tail regions in a protective cocoon, shielding them from the solvent. Driven

by the hydrophobic effect, increasing surfactant concentration, as well as altering the head-to-tail size ratio, provides a wide assortment of liquid crystalline phases, such as cylindrical micelles, bilayers, or inverted micelles. Two examples of such mesophase structures are shown in Fig 1.3.[18–20]

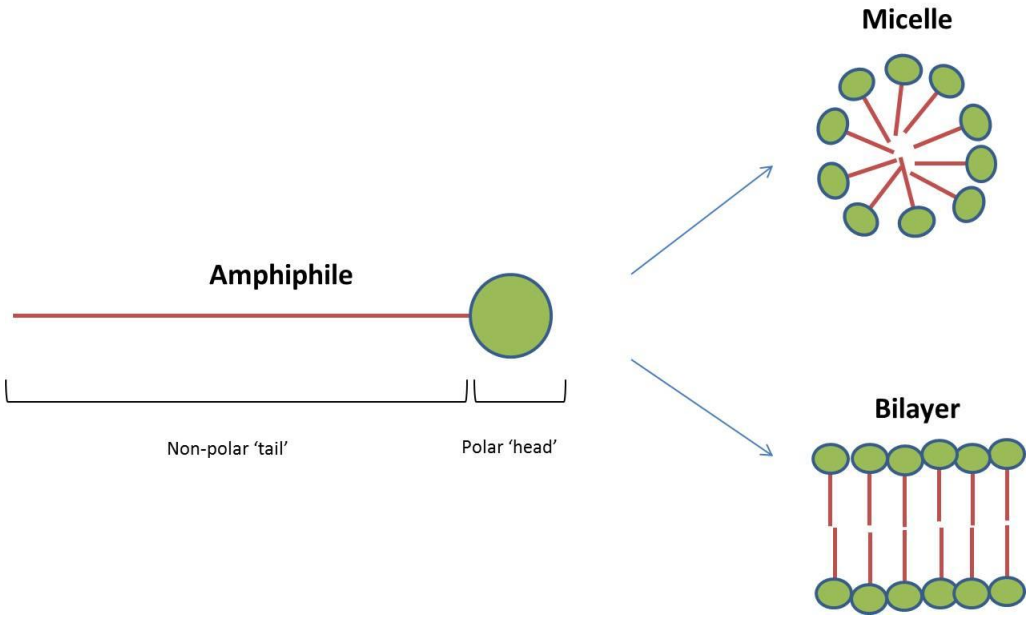

Figure 1.3: General structure of an amphiphile (left), two examples of mesophase structures formed at low and high concentration (right).

Chromonic mesogens, by contrast, have not only a distinctly different general molecular structure, but also exceptionally different behaviour in solvent, as discussed next.

1.2 Chromonic liquid crystals

1.2.1 Molecular structure and aggregation in solution

The second type of lyotropic liquid crystal is termed chromonic,[1] as exhibited by a range of nucleic acids, dyes and even some carcinogens. Chromonic mesogens generally have a similar molecular structure: a rigid-flat-extended aromatic core, with ionic or hydrophilic groups attached to the periphery to provide solubility in a polar solvent. In aqueous solvent, enthalpic interactions typically drive face-to-face aggregation of chromonic molecules, forming stacks in solution, with an average

spacing of 0.34 nm. These can be visualised as being like many stacks of pound coins randomly orientated in a solution, as shown in fig 1.4.

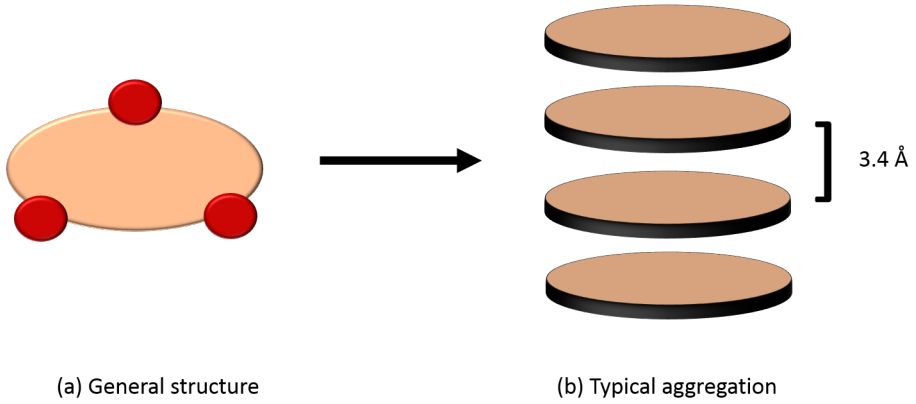

Figure 1.4: General molecular structure of a chromonic mesogen (left) with a flat, rigid aromatic core (peach) and periphery solubilising groups (red) and typical stacking behaviour in solution (right).

1.2.2 Structure within an aggregate

Chromonic aggregates are typically composed of mesogens that associate in a face-to-face style aggregation.[1,21] In spite of the seemingly simple stacking mode of aggregation, a wide range of structures within the column itself are possible. Subtle differences in the molecular structure of the mesogen alter how the molecules interact on self-assembly (whether due to steric or charge interactions) and thus the overall structure within a column. A range of possible structures can be seen in fig 1.5.

In addition to the typical face-to-face self-assembly of chromonics, more exotic (and in some cases controversial) structures have been proposed in the literature.[5,22] One common structure suggested is the *double-width column*. The structure possesses two molecules in each 'layer' of the column, whereby the neighbouring 'layer' within the stack is two molecules rotated by 90°.

This, as well as larger and more complex architectures, such as the hollow chimney structure,[23,24] and the *brickwork layer structure*[3,25] have been proposed (fig 1.6).

The double-width column and the brickwork layer structure will be discussed in detail in chapter 3, 4 and 5.

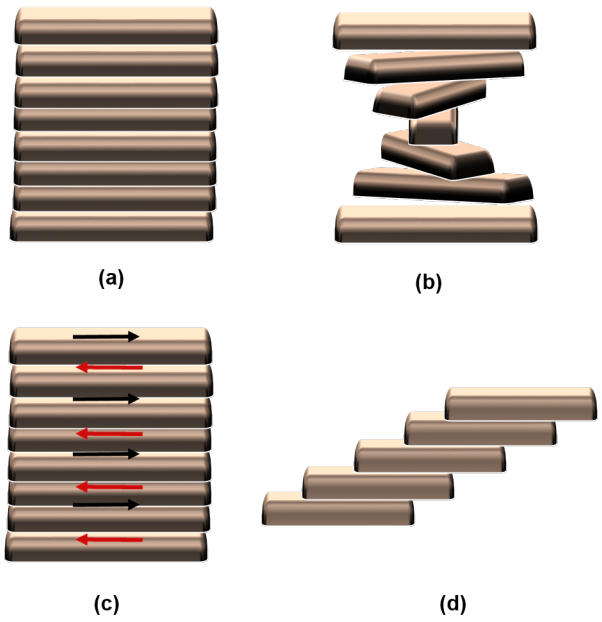

Figure 1.5: Structures within a column, typical face-to-face aggregation of H-aggregates (a), Chiral stacking (b), anti-parallel arrangement (c), offset stacking of J-aggregates (d).

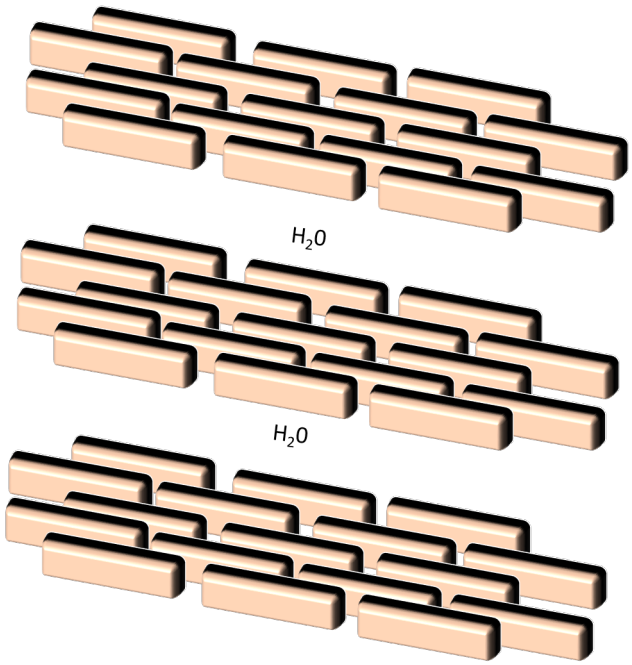

Figure 1.6: Schematic diagram showing the brickwork layer structure.

1.2.3 Formation of chromonic mesophases

In dilute solution, the aggregation of chromonic mesogens leads to a distribution of stacks of various sizes. At low concentration, when the motion of these stacks are largely uncorrelated with respect to one another, the phase is liquid and is usually termed termed the isotropic phase (I). Increasing concentration of the mesogen decreases free volume in the system, forcing parallel arrangement of aggregates, giving rise to a liquid crystal phase. Lower concentrations in the nematic (N) phase, where stacks are on average pointing in the same direction (long range orientational order, short range translational order). Higher concentrations yield a hexagonal phase (M), where the columns possess statistical hexagonal symmetry (where centers of columns lie on a hexagonal lattice), thereby providing not only orientational order, but long range translational order. These structures can be seen in fig 1.7. In contrast to thermotropic phases, where the director is defined with respect to individual mesogen orientations, the director of a chromonic system is defined with respect to the aggregate as a unit.

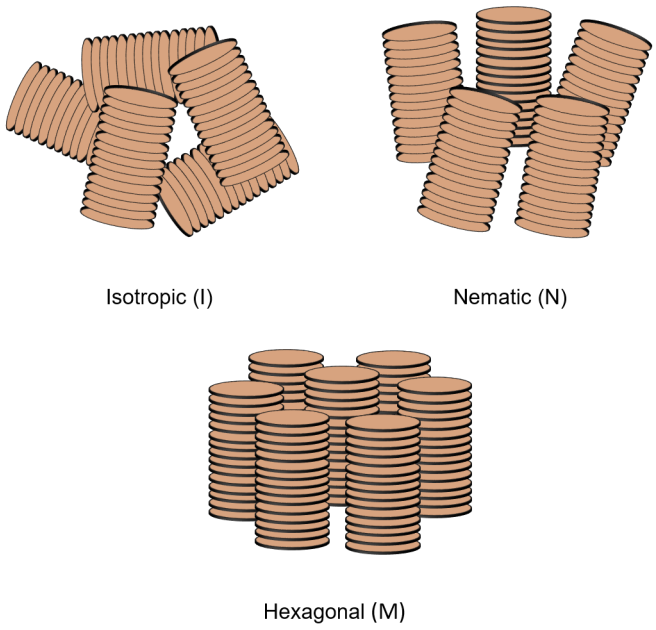

Figure 1.7: Typical structure of chromonic mesophases.

Phase diagrams are frequently used to show how the state of the system changes as a function of concentration of mesogen and temperature. Phase diagrams are essentially a map, with each region of the map corresponding to a different physical

state for the system. The phase diagram of the chromonic system DSCG/water (which shows the standard pattern for chromonic systems) can be seen in fig 1.8. Unlabelled regions correspond to two phase regions separating different mesophases.

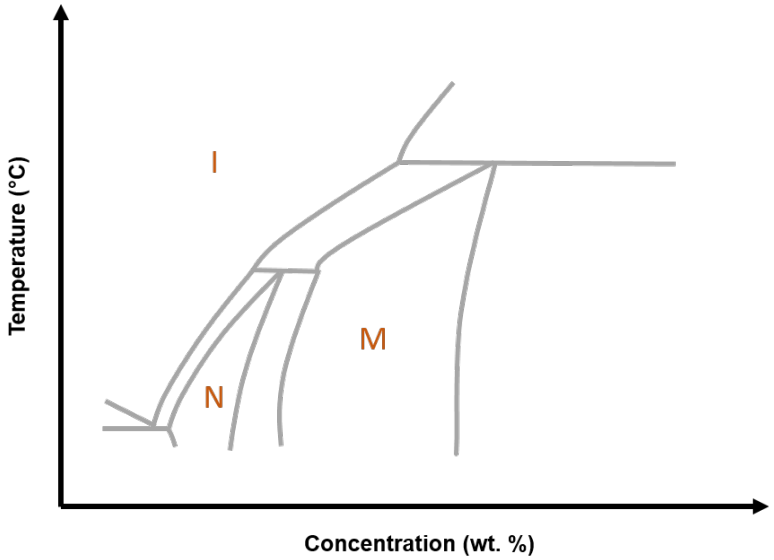

Figure 1.8: Typical chromonic phase diagram for the DSCG/water system. Taken from Lydon.[1]

1.2.4 Onsager theory

A simple model used for lyotropic liquid crystals to predict the isotropic-nematic phase transitions is the Onsager hard-rod model.[26] The theory represents the molecule (or linear stack of molecules) as an idealised cylinder, so that the system can be viewed as a 3-dimensional gas of hard rods. In this system, there are two components to the entropy: translational and orientational. Both of these entropic components are coupled, and the favoured component is dependent on the density of the system.

At low density, orientational freedom is favoured, due to the negligible gain in entropy made by reducing the excluded volume from a parallel arrangement of rods. At higher densities, translational freedom from a more aligned system is more favourable, with the loss of orientational entropy now occurring. This transition from a disordered to an ordered arrangement is the isotropic-nematic transition.[27,28]

1.2.5 Comparing chromonics and conventional amphiphiles

Chromonic and conventional amphiphilic systems, both categorised as lyotropic liquid crystals, have extensive differences in molecular structure, leading to extremely different behaviour and physical properties in solution. Chromonic systems have no critical micelle concentration (CMC) and aggregate even in dilute solutions, whereas conventional amphiphiles will have a specific concentration above which micelles form. The aggregation process in a chromonic system is isodesmic, where each new addition of a molecule to a stack has the same energy increment. In contrast, conventional amphiphiles have a non-isodesmic association, where the energy increment of adding a molecule is governed by the size of the micelles. The driving force for aggregation, predominantly enthalpic in chromonic systems and entropic in amphiphilic systems, results in very different phase diagrams. The progressive formation and growth of chromonic stacks as a function of concentration leads to a multi-peritectic phase diagram. In contrast, the conventional amphiphilic system has an extensive range of micelle types as a function of concentration, leading to a multi-eutectic form.[1,17]

1.3 Applications of chromonic systems

1.3.1 Widespread recognition of potential

Despite research as early as 1915 showing a chromonic nematic phase in the aqueous solution of phenanthrene sulfonic acid,[29] and research in the 1920's showing chromonic mesophase formation by the medicinal drugs salvarsan and neosalvarsan,[30] it is only in the last twenty to thirty years that the term chromonic has been more widely accepted as a type of lyotropic liquid crystal phase. Lyotropic chromonic liquid crystals have a unique set of properties and characteristics that have already been exploited with promising results in biosensors, polarizing films, and micro-patterned materials, with further details for two specific applications detailed next.[1,31–37]

1.3.2 Photovoltaic device

With the majority of the world's energy demand being suppled by fossil and nuclear sources, key issues such as non-sustainability and environmental damage caused by pollution, provide a strong drive for the development of cleaner and more renewable energy sources. The advancement of photovoltaic and photo-electrochemical cells is a key area of current research. The harvesting and conversion of light energy into electricity has the potential to reshape the energy industry. However, the low conversion efficiency of current light-harvesting systems makes this energy alternative a somewhat minor contributor to the global energy supply.[38]

In principle, the operation of a solar cell is comparable to a simple semiconductor device that converts light energy into electricity. Firstly, light shines onto the crystal atoms embedded in the solar cell. Free electrons are created, due to ionization of the crystal atoms, creating an electron-hole pair. The solar cell works to separate this pair, where a current is produced by the movement of the electrons, therefore attempting to reduce recombination of the electron-hole pair.

Chromonic liquid crystals possess unique electro-optical properties, as the strong electron accepting properties and high conductivity of the pi-pi interactions allows for efficient one-dimensional charge transfer and energy transfer. Incorporating chromonic aggregates into a photovoltaic device allows them to act as a secondary electron acceptor, thus decreasing the back flow of electrons, and therefore decreasing electron-hole recombination. A film of chromonic stacks are deposited onto the solid substrate (e.g. where an electron rich group is attached to a fullerene, such as a porphyrin-fullerene dyad layer) of the photovoltaic device. The chromonic aggregates work by transferring energy from the photon (from sunlight), to the fullerene network below. Charge separation between the porphyrin-fullerene dyad then directs an electron up the chromonic stack, where it is transported to a cathode. An example set-up, using the chromonic mesogen perylenediimide (PDI) derivative PDI12, is shown in fig 1.9.[39]

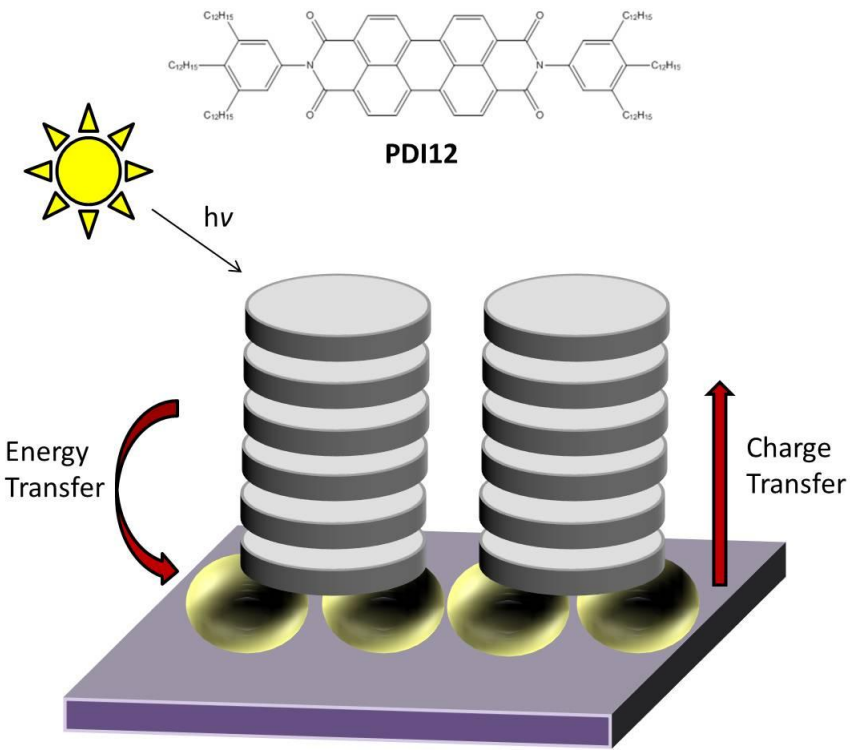

Figure 1.9: PDI12 molecular structure and schematic of chromonic columns integrated into a photovoltaic device.

1.3.3 Templates

One of the key frontiers in modern materials chemistry involves using molecular self-assembly to construct devices and materials, with one example being the micro- and nano-patterning of directionally-dependent materials. Chromonic liquid crystals provide a unique set of self-assembly and alignment properties, which when incorporated on a template, can provide an efficient and inexpensive way of producing aligned solid dye films. The fabrication process entails three main parts. The initial stage requires the production of a micro-grooved template (a strip of polymer, such as polydimethyl siloxane, with arrays of parallel microgrooves on the surface). Introduction of an isotropic solution of the perylene di-carboximide chromonic dye to the template (encased by a cover slip). Drying by peripheral evaporation, allows the formation of aligned M phase chromonic ribbons to form in each micro-groove. The re-usable polymer template can then be peeled off, leaving an anisotropic solid chromonic dye on the cover slip. This process is shown in fig 1.10.[31,40]

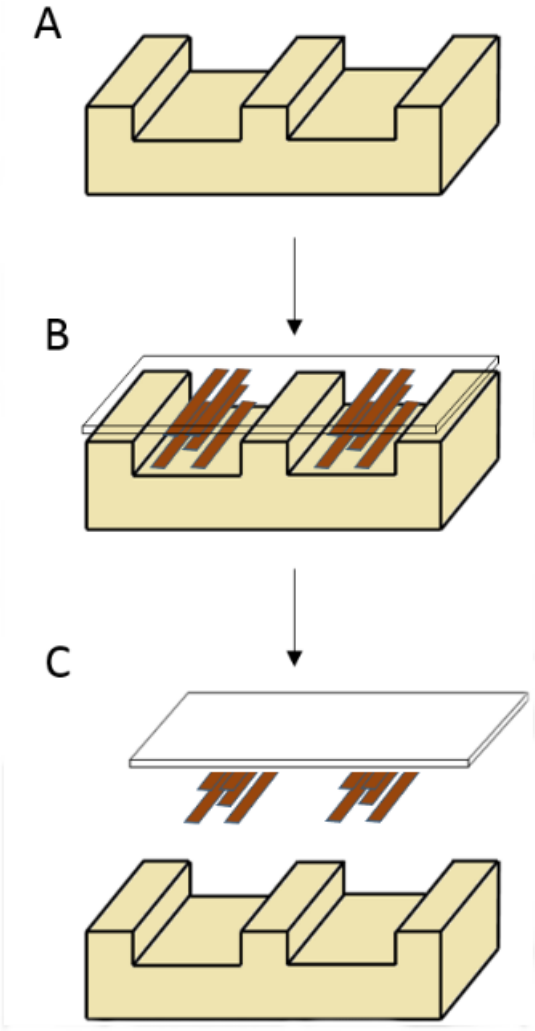

Figure 1.10: Creating aligned solid dye on a film using a template.

1.4 Overview and scope of thesis

The aim of this thesis is to use computational methods to successfully model the behaviour of ionic chromonic mesogens in solution. Different techniques are employed to attempt to tackle a range of different time and length scales, allowing for both the aggregation mode within each column, as well as the larger scale interactions between columns.

The aggregation and phase behaviour of a chromonic system is highly sensitive to the molecular structure of each mesogen, with small changes in structure able to produce a vast array of different aggregates. The aims of this work is to simulate a range of mesogens, and identify possible design rules based on the different packing behaviours observed. Using these design rules to successfully predict stacking behaviour, based on the molecular structure, would revolutionise the level of control the user would have over the chemical system.

Chapter 3 and 4 focus on atomistic models of 3 specific chromonic dyes using molecular dynamics, in which all interactions are accounted for (including long range electrostatic interactions). Experimental literature proposes several exotic aggregate structures for these dyes, sparking interest that a computational model could be developed to match key experimental results (cross sectional areas, x-ray diffraction patterns, stacking free energy change), in an attempt to either add credence to these exotic structures, or reinterpret the results in terms of alternative structures.

In chapter 5, the possibility of using larger scale simulations with coarse-grained models (where many atoms are grouped into beads to reduce the number of sites) is investigated. Smaller test systems are initially modelled, and then the larger, more complex, chromonic dyes, with the ambitious aim of simulating a whole mesophase. Comparison of different coarse graining techniques (for specific use on chromonic systems) are discussed, as well as a comparability of key structural properties (radial distribution functions, stacking distances) between the coarse grained and atomistic simulations, allowing for the validity of the model to be probed.

An in-depth look at the theory behind the computational methods employed, with a heavy focus on atomistic simulations and details for analysis techniques used during research, are provided in chapter 2. Any additional theory is provided in the

relevant research chapter when required (chapter 5 has an extensive introduction to coarse-graining techniques).

Finally, chapter 6 provides a concise summary of all the key findings presented in each of research chapters, as well as future avenues of research.

Chapter 2

Computational methodology

2.1 Introduction

Theory has always played a vital yet challenging role within the realm of chemistry, with laws and equations being derived to provide a mathematical description for the behaviour of molecules.[41] A significant weakness of theory is that even for simple systems, equations typically evolve into highly complex mathematical expressions that cannot be solved analytically. Computer simulations bridge the gap between theory and experiment by providing essentially exact results for a model of the system of interest. Direct comparison between the simulation and experimental results allows for the correctness of the underlying model to be evaluated. Successful matching of key thermodynamic or structural data provides not only reassurance in the validity of the model used, but can offer interpretation of new results for the experimentalist. Simulations can reveal a molecular level picture of a chemical system, at a level of detail typically inaccessible to the experimentalist, thus providing a link between the microscopic world (molecular geometry, short range interactions between individual molecules) and the macroscopic world (long range aggregation behaviour, phase transitions). Simulations also have several advantageous features from a practical perspective. Exploration of systems at conditions unfavourable for an experimentalist, such as extremely high temperatures or pressures, can be safely achieved on a computer. Construction of any type of molecule, whether known or

not, is also done with relative ease compared to the timescale and expense required for a very complicated synthesis process.[42]

2.2 Simulations of chromonic systems

Simulations can be used as a powerful tool for research into chromonic liquid crystal systems. Previous experimental studies, using optical microscopy, multi-nuclear NMR and X-ray diffraction, were employed on Sunset Yellow (SSY), an anionic azo dye commonly used as a food dye.[43–46] The mesogens were found to stack into aggregates, with the cross sectional area of these aggregates being deduced to be one molecule thick (typical H-aggregates). A molecular simulation study of SSY in aqueous solution observed the aggregation of dye molecules into stacks.[47] Key thermodynamical and structural features from experiment, such as the binding energy and the intermolecular distance between dye molecules in a stack, were matched. Confident in the correctness of the underlying model, the simulation results were also able to provide a molecular level picture of the stacking arrangement of dyes within a column, with neighbouring dye molecules showing a head-to-tail stacking preference, and the direction vector of the carbonyl group alternating between neighbour molecules in the stack.

TP6EO2M is a non-ionic chromonic designed to control anchoring of a liquid crystal molecule to a surface. Experimental studies using proton NMR,[48–52] X-ray diffraction,[48,49] and DSC[50] show that TP6EO2M molecules exhibit a strong desire to aggregate in a face-to-face manner in aqueous solution, with a relatively large interlayer distance (distance between neighbouring dye molecules in an aggregate) than seen in many typical chromonic systems. Atomistic simulations[53] were able to successfully match not only the stacking behaviour found in experiments, but also the stacking free energy change, demonstrating not only a structurally accurate model, but also a model able to capture the essential thermodynamics of the system. The molecular level detail provided in the simulation revealed that the larger-than-average interlayer spacing was attributable to the repulsion between the

side chains present in the molecule, a feature extremely difficult to distinguish using experimental techniques.

2.3 Atomistic models

When attempting to tackle a problem using molecular modelling, the size and complexity of the system is central in determining the level of simulation needed. Quantum mechanical methods involve considering the electrons explicitly. Electron interaction is obtained from an approximate solution of the electronic Schrödinger equation and the potential energy function as a sum of the nuclear repulsion. Not only is this method extremely accurate, but it also allows properties to be calculated that are dependent upon the electronic distribution in a molecule (such as electron transfer and bond breaking/making). Despite these benefits, electronic structure calculations are limited to very small systems (a few hundreds of atoms). The complexity of the calculation needed to model a small system of molecules (which can easily be of the order of 50,000 atoms) is effectively impossible to simulate using this method. Molecular mechanics force fields, on the other hand, remove electronic motion and calculate the potential energy as a function of the positions of the nuclei only. Molecules can be represented by discrete atomic sites, with chemical bonds between these sites being represented by springs. This dramatically reduces the size of the calculation, thus allowing the system size of interest to be significantly larger than with quantum mechanical methods. It is for these reasons that the interactions between particles of liquid crystal systems are predominantly modelled using a force field approach.[41,42,54]

2.4 Molecular mechanics

Molecular mechanics is based on a fairly simple and well-characterised model, where each molecule is seen as a collection of atoms. Interactions between atoms of the same molecule are modelled with bond stretching, angle bending and torsional potentials, and interactions between different molecules represented by electrostatic

and van der Waals potentials. Each term and component of a force field will be described in detail next.

2.4.1 Force fields

In molecular modelling, the force field, also commonly (and more correctly) referred to as the potential, is the mathematical functions and parameters used to calculate the potential energy between all atoms or particles within a system. Contrary to a quantum mechanical approach, a fully classical potential treats the electrons implicitly, by having the atoms or particles move along a potential energy surface. The force field parameters can be found from either experimental data, theoretical data or higher level quantum mechanical calculations. In a typical molecular mechanics simulation, the two energy contributions to the potential are the intramolecular interactions, between atoms within a molecule, and the intermolecular interactions between molecules in the system.[41]

$$E_{\text{ff}} = E_{\text{bonded}} + E_{\text{non-bonded}} \tag{2.1}$$

where E_{ff} is the total potential energy, E_{bonded} is the energy contribution from bonded interactions within a molecule in the system and $E_{\text{non-bonded}}$ is the energy contribution from non-bonded interactions between molecules in the system. The components of both bonded and non-bonded interactions can be broken down into a series of discrete interactions which, upon summation, give a total energy of the molecule in the system.

$$E_{\text{bonded}} = \sum E_{\text{bonds}} + \sum E_{\text{angles}} + \sum E_{\text{dihedrals}} \tag{2.2}$$

$$E_{\text{non-bonded}} = \sum E_{\text{elec}} + \sum E_{\text{vdw}} \tag{2.3}$$

$$E_{\text{ff}} = \sum E_{\text{bonds}} + \sum E_{\text{angles}} + \sum E_{\text{dihedrals}} + \sum E_{\text{elec}} + \sum E_{\text{vdw}} \tag{2.4}$$

The bonded interactions within a molecule in the system can be broken down into three key interactions: bond stretching, angle bending, and dihedrals. The non-bonded interactions between non-bonded atoms and molecules in the system can

be broken down into long range electrostatic and van der Waals terms. All of the above interactions are visually represented in fig 2.1. Each interaction will now be expanded on in the following section.

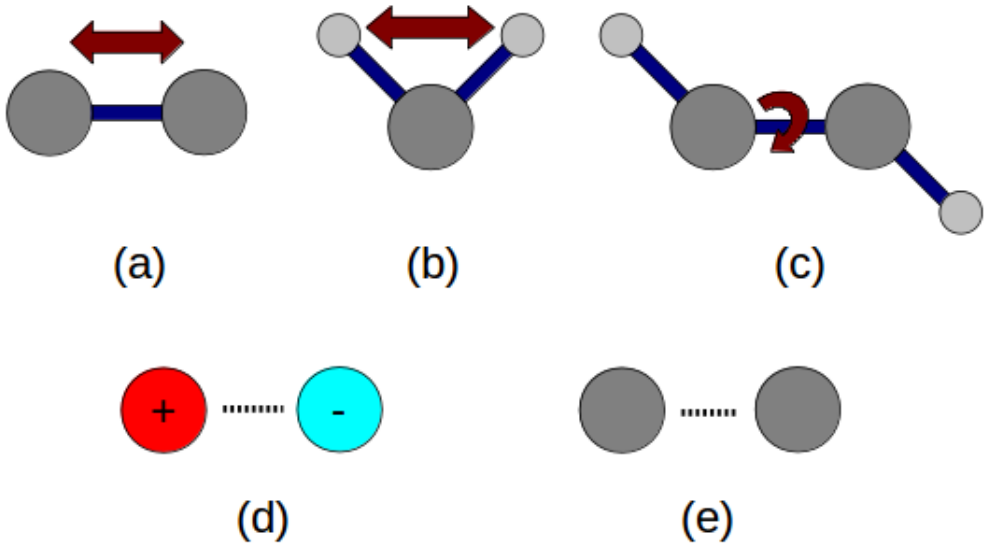

Figure 2.1: Visual representation of the bonded and non-bonded interactions included in a force field. (a) bond stretching, (b) angle bending, (c) dihedral, (d) electrostatic forces, (e) van der Waals forces.

2.4.2 Bonded interactions

Bond stretching

The bond stretching term describes how the energy between two atoms in a molecule changes as a function of the length of that bond connecting them. A common way to express the energy is to apply the Taylor expansion. Truncating after the second-order term is known as the harmonic approximation, giving rise to the following equation:

$$E(r) = \frac{1}{2}k(r - r_0)^2, \tag{2.5}$$

where r_0 is the reference bond length, r is the actual bond length, and k is the harmonic force constant given by

$$k_r = \frac{d^2E}{dr^2} \mid_{r=r_0} . \tag{2.6}$$

The bond length corresponding to the lowest energy between two atoms is termed the reference or equilibrium bond length. As the two atoms move in closer proximity to each other, this leads to a very rapid increase in energy, mimicking real life systems whereby the electron clouds on each respective atom would overlap and experience a large repulsive force. Similarly, as the bond length is stretched too far from the equilibrium bond length, a rapid increase in energy occurs. It should be noted that within a molecular mechanics approach, the bond cannot break. Despite the harmonic approximation being justifiable only for small deviations from the reference bond length, the magnitude of covalent bond vibrations are within this interval of distances, meaning it is unnecessary to describe deviations larger than this interval (as shown in fig 2.2). In particular cases where accuracy is needed for a far broader range of distances, the Morse potential is particularly good at representing the potential energy for a bond stretch

$$V(r) = D(1 - \exp^{-a(r-r_o)})^2, \tag{2.7}$$

where D is the well depth, a is the width of the potential, described as

$$a = \sqrt{\frac{K}{2D}}. \tag{2.8}$$

where K is the force constant at the minimum of the well. Whilst the Morse potential offers a more realistic description of a covalent bond stretching, the presence of the exponential in the function makes the use of this potential extremely computationally expensive. For these practical reasons, the harmonic approximation is the common choice in most molecular mechanics studies.[41,42]

Angle bending

The angle bending term describes how the energy between three connected atoms in a molecule changes as a function of the angle between them. A common approach is to use a harmonic potential

$$E(\theta) = \frac{1}{2}K_\theta(\theta - \theta_0)^2, \tag{2.9}$$

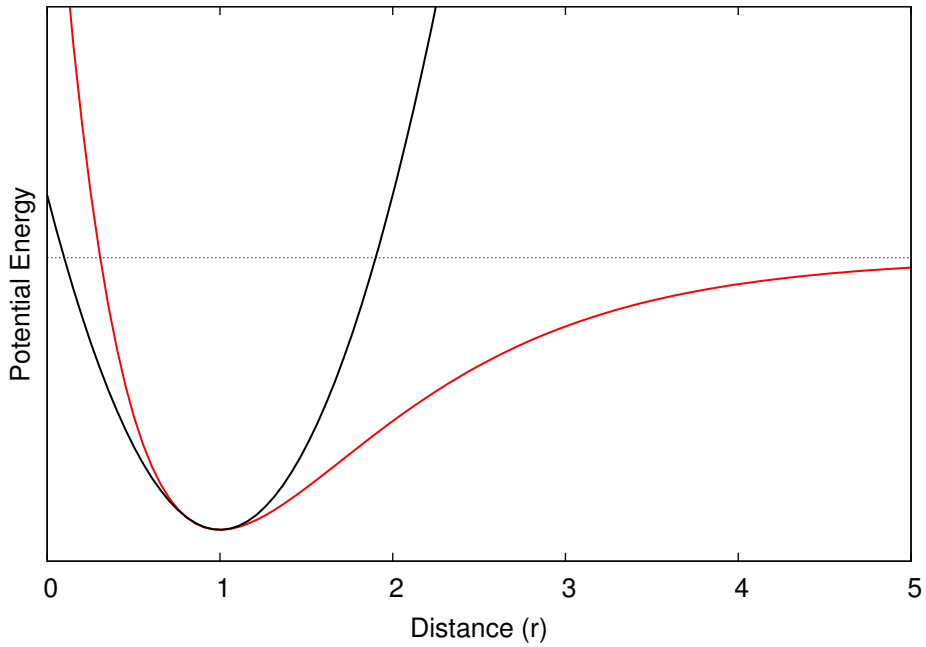

Figure 2.2: A comparison of the Morse Potential (red) and the harmonic approximation (black) for the bond stretching term. A zero energy line has been added (dotted).

where K_θ is the angle bending force constant, θ is the actual angle and θ_0 is the reference angle. The angle corresponding to the lowest energy between three connected atoms is termed the equilibrium angle. As with bond stretching described above, a deviation from this equilibrium angle causes an increase in energy, albeit smaller than bond stretching deviations.[41,42]

Dihedral angle

The dihedral angle, also commonly referred to as the torsional angle, describes the angle between two intersecting planes produced by four connected atoms (where two central atoms plus either end atom give rise to two planes). The dihedral angle term therefore is how the energy changes as a function of the dihedral angle between these two planes

$$E(\omega) = \sum_n \frac{1}{2} V_n \cos(n\omega), \tag{2.10}$$

where V_n are the rotational force constants and ω is the actual dihedral angle.

In comparison to both bond stretching and angle bending, large deviations in

dihedral angle are far more prevalent, due to the internal rotation barriers being low. The dihedral term is particularly important for long molecules, such as polymers and/or liquid crystal systems, where molecular flexibility can have a large effect on how the molecules interact within the system.[41,42]

2.4.3 Non-bonded interactions

Electrostatic forces

Electrostatic forces arise from the unequal spread of charge in a molecule, due to the presence of permanently electron-rich or electron-poor atoms or regions. Observing a classic example of H-Cl, we have an electronegative chlorine, due to the property of attracting electron density towards itself, leaving an electropositive hydrogen atom. Modelling this unequal distribution of charge can be achieved by placing point charges on each of the atomic sites in the molecule. A Coulomb potential energy function can then be solved to model the energy of these partial charges as a function of distance from one another

$$E_{\text{elec}} = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r_{ij}}, \tag{2.11}$$

where ϵ_0 is the permitivitty of free space, q_i and q_j are atomic charges for atom i and j, and r_{ij} is the distance between atom i and j.

Long range electrostatic interactions

A problem with the electrostatic potential is that it is extremely long ranged, with the imposed periodic boundary conditions on the simulation cell causing essentially infinitely ranged forces. A method to deal with this is to partition the interactions into two categories: short and long ranged forces. Short ranged forces can be swiftly dealt with by applying a cut-off (r_c) to the potential, where beyond the cut-off distance, the potential is set to zero. For the long range forces, the central simulation cell and how each particle interacts with periodic images, results in an edited coulombic potential as shown in Eq 2.12, where nL is the lattice vectors. A key problem is that this function not only converges extremely slowly, but is fully

dependent on the order by which the terms are considered. An Ewald sum, by which the summation is first converted into two series that converge far more rapidly and efficiently, is employed to overcome this issue.[55]

In the Ewald method, two stages are implemented. The first stage is where each point charge in the system is surrounded by a charge distribution of equal magnitude but opposite sign, thus neutralizing the original charge (summed in real space). In the second stage, a canceling charge is added (summed in reciprocal space)

$$E = \frac{1}{4\pi\varepsilon_0}\frac{1}{2}\sum_{n}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{q_i q_j}{r_{ij}+nL}. \tag{2.12}$$

A cheaper alternative is to use a Particle-Particle-Particle-Mesh (PPPM), whereby continuous coordinate point charges are replaced by a grid-based charge distribution. The discrete Poisson equation can then be solved for this grid to generate the potential and forces. Other alternatives, not discussed here, include the reaction field method[56] and the cell multipole method.[42,57]

Van der Waals interactions

Van der Waals forces are used to encompass a variety of interactions between molecules in a system not included in the electrostatic interaction. The most common interactions described by the Van der Waals force include dispersion, repulsion and induction. Dispersion describes the attraction between molecules, attributable to the motion of electrons in different atoms correlating, leading to a lowering of the energy. Repulsion represents the force experienced when atoms are in close proximity to one another, where the overlapping of negatively charged electron clouds leads to a strong increase in energy, thereby moving further apart to lower the energy. Finally, induction depicts the change in charge distribution that arises due to an atom or molecule nearby distorting the electron distribution on another atom or molecule. All of these interactions, amongst others, can be modelled by the very common Lennard-Jones potential, graphically represented in fig 2.3

$$V_{LJ}(r) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right], \tag{2.13}$$

where ϵ_{ij} is the well depth of the potential, and σ_{ij} is the distance at which $V_{LJ} = 0$.

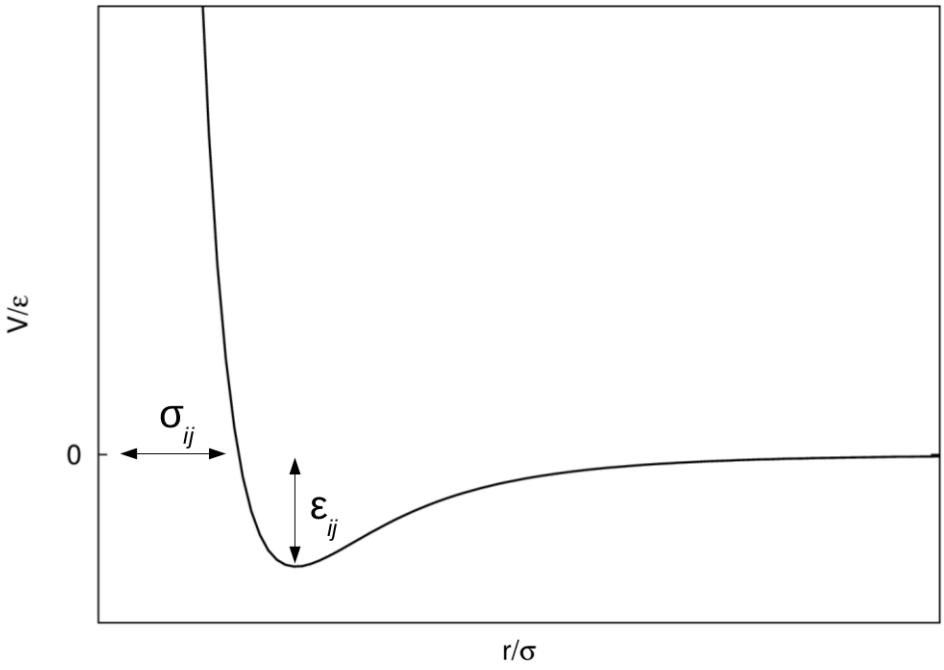

Figure 2.3: Lennard-Jones potential commonly used to model the van der Waals interactions between molecules in a system.

Choosing a force field

Selecting a suitable force field for a system is crucial for generating successful results. A good choice of force field requires detailed knowledge of the molecule of interest, the available force fields and how they were parameterised, as well as any literature studies on the success of similar molecules simulated.

For condensed matter systems, such as proteins or polymers, a range of force fields are extremely popular, such as OPLS,[58] GROMOS[59] and GAFF.[60] For all of these force fields, the interaction potentials for bond stretching and bending are typically restricted to a harmonic approximation, and a Lennard-Jones potential used to model the Van der Waals interactions. Certain force fields, such as GROMOS and GAFF were parameterised specifically for biological systems, such as nucleic acids.

2.4.4 Molecular dynamics

Molecular dynamics (MD) is a simulation technique used to study the average behaviour of a many-particle system, by following the natural evolution of the system as a function of time. The nuclear motion of the constituent particles obey the laws of classical mechanics, meaning that for an N body system interacting through potential functions, consecutive configurations of the system can be generated by using Newton's second law of motion. The acceleration term can be written as a second-order differential equation using positions of particles as a function of time, to give the expression

$$\mathbf{F}_i = m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2}, \tag{2.14}$$

where $\mathbf{F}_i$ is the force of the i th particle and m is the mass of the i th particle. The force on each particle (due to the interaction between all of its neighbours) can be calculated as a function of the potential energy, allowing the second-order differential equation in 2.14 to be solved

$$\mathbf{F}_i = -\nabla_i V(\mathbf{r}_1, ..., \mathbf{r}_N). \tag{2.15}$$

As the potential energy functions are coupled together, a many body problem arises which cannot be solved analytically (with a pen and paper), introducing the need for a numerical approach. The finite difference method allows for the equations of motion to be solved on a step-by-step basis, by allowing time to be fragmented into discrete time segments δt. Using the knowledge of the positions of all the particles at present time t and the previous time step, combined with the force acting on each particle given to us by the relationship to the potential energy, allows us to predict the positions at the next time step. This allows the equations of motions to be solved for a many body problem, generating successive configurations and thus generating a simulation of the behaviour of the particles as a function of time.[41,42,54]

Integration algorithms

Using the finite difference method, where time is discretized, to evaluate the equations of motion in steps, requires the use of an integration algorithm. The three

most common integration algorithms are the Verlet algorithm,[61] the leap-frog algorithm,[62] and the velocity-Verlet algorithm.[63] The simplest of the algorithms is the Verlet algorithm. The finite difference method allows time to be broken down into discrete time lengths δt, with the next successive moment in time being written as $t + \delta t$, $t + 2\delta t$ etc as the system progresses with a particular time step. A Taylor expansion series can be written for the positions about time t, in both the forward and backwards direction.

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 \tag{2.16}$$

$$\mathbf{r}(t - \delta t) = \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 \tag{2.17}$$

The summation of these giving a time-reversible expression for the position at the next step

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \mathbf{a}(t)\delta t^2, \tag{2.18}$$

where the velocity can then be calculated independently from:

$$\mathbf{V}(t) = \frac{\mathbf{r}(t + \delta t) - \mathbf{r}(t - \delta t)}{2\delta t}. \tag{2.19}$$

Whilst the Verlet algorithm provides good energy conservation, as well as several advantageous practicalities such as simplicity of code, the summation may introduce numerical imprecisions, and the velocities are treated poorly. Computing the trajectory does not implicitly require a velocity. However, velocity is needed for the estimation of the kinetic energy of the system, which in turn is related to the total energy, temperature and pressure of the system, and thus is extremely important.

The development of the leapfrog algorithm[62] allowed for several of the Verlet algorithm shortcomings to be dealt with.

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \mathbf{v}(t + \frac{1}{2}\delta t)\delta t \tag{2.20}$$

$$\mathbf{v}(t + \frac{1}{2}\delta t) = \mathbf{v}(t - \frac{1}{2}\delta t) + \mathbf{a}(t) \tag{2.21}$$

Here, the mid-step velocities are calculated and then used to calculate the new positions (velocity leaps over the positions, which in turn leaps over the velocities). All of this information can then be used to calculate the velocity at present time t,

$$\mathbf{v}(t) = \frac{1}{2}(\mathbf{v}(t + \frac{1}{2}\delta t) + \mathbf{v}(t - \frac{1}{2}\delta t)). \tag{2.22}$$

Despite the leapfrog and Verlet algorithm being algebraically equivalent, the leapfrog is a slight improvement upon the Verlet due to the increased numerical precision. One of the drawbacks of the Verlet algorithm is that the positions $\mathbf{r}(t+\delta t)$ are obtained by adding a smaller term $\delta t^2\mathbf{a}(t)$ to the difference of the two much larger terms $2\mathbf{r}(t)$ and $\mathbf{r}(t-\delta t)$, typically leading to a loss of precision. At no point does the leapfrog algorithm take two large numbers away from each other to give a smaller one, thus minimises the loss of precision.[41,42]

The last alternative is the Velocity Verlet algorithm,[63] a method where positions, velocities and accelerations are stored at the same t. While the Velocity Verlet algorithm provides a more accurate trajectory than the previously mentioned algorithms, this comes with an increase in computational cost. The leap-frog algorithm, despite its several shortcomings, is commonly the preferred choice for molecular simulations for its overall simplicity and speed.

2.4.5 Molecular dynamics summary flowchart

A simple flowchart giving a brief overview of a molecular dynamics simulation process can be seen in fig 2.4.

Periodic boundary conditions

One of the biggest weaknesses in computational modelling is simply the system size that we can explore. The system size can impose limits on not only the phenomena that the user is trying to simulate, but also (on a fundamental modelling level) how isotropic the force experienced is throughout the system. Molecules in the centre of the system have a markedly different environment to the molecules at the surface. This leads to an unequal, and potentially unrealistic, description of the forces acting

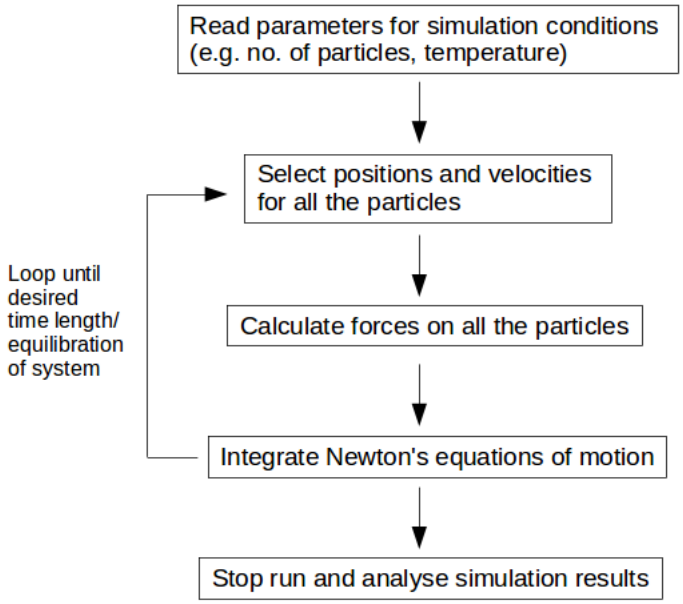

Figure 2.4: Flowchart of a molecular dynamics simulation.

upon that molecule, dependent on its position in the system. Fortunately, surfaces are easily eliminated from a simulation by employing periodic boundary conditions. The initial box is surrounded by an infinite number of identical systems, allowing for each molecule to be simulated in the bulk. To increase computational efficiency by decreasing the level of calculation needed, short ranged forces are truncated (with a cut-off distance less than or equal to half the box length), also allowing no atom to interact with its periodic image. A two dimensional illustration of periodic boundary conditions can be seen in Fig 2.5.

Thermodynamic ensembles

Microscopic simulations are used in an attempt to mimic and give insight into macroscopic systems. A thermodynamic ensemble is required to be able to effectively compare the two systems. Within a particular thermodynamic ensemble, the simulation is performed with a particular set of fixed state variables. Each ensemble should lead to the same macroscopic thermodynamic state. There are a wide variety of thermodynamic ensembles that can be used in simulation:

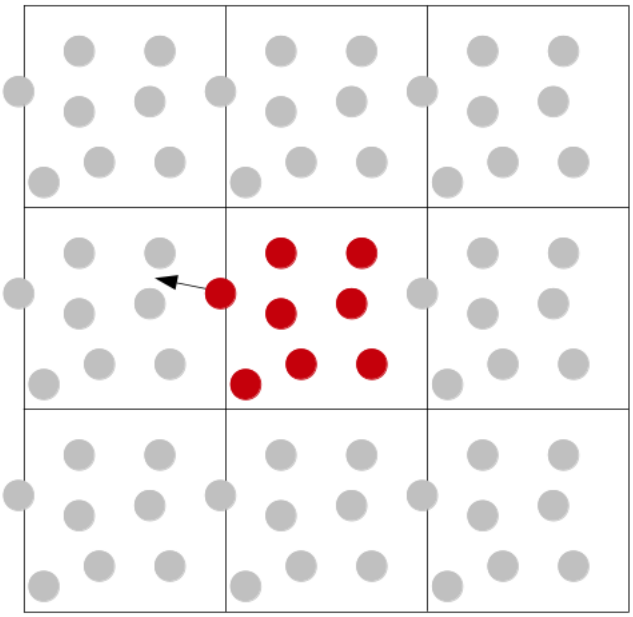

Figure 2.5: A two dimensional representation of periodic boundary conditions, where the central box containing red atoms is the simulation cell, and the surrounding grey atom systems are periodic images of the original cell.

- **Microcanonical**: Fixed number of atoms, N, volume, V, and energy, E
- **Canonical**: Fixed number of atoms, N, volume, V, and temperature, T
- **Grand canonical**: Fixed chemical potential, μ, volume, V, and temperature, T
- **Isothermal-isobaric**: Fixed number of atoms, N, pressure, P, and temperature, T
- **Isoenthalpic-isobaric**: Fixed number of atoms, N, enthalpy, H, and pressure, P.

Newton's equations of motion provide a default microcanonical ensemble. However, in an attempt to mimic the conditions of real experiments, it is typical to use an isothermal-isobaric ensemble for many simulations.

Temperature Coupling

The temperature of the system can be readily controlled, due to the relationship of temperature and the average kinetic energy, which in turn can be controlled by scaling velocities. A well-known thermostat to use is the Berendsen thermostat,[64] where the system is weakly coupled to an external heat bath at a given reference temperature T_0. Any temperature deviation from T_0 will result in the system being slowly corrected to the reference temperature.

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \tag{2.23}$$

where τ is a time constant. The main disadvantage to employing a weak-coupling scheme is that, whilst it is efficient in relaxing a system to the reference temperature, it suppresses the fluctuations of the kinetic energy, therefore failing to generate a true canonical ensemble. The development of the Nosé-Hoover algorithm,[65] first proposed by Nosé[66] and further modified by Hoover,[67] where a thermal reservoir and friction term are introduced into the equations of motion, thereby ensuring a correct canonical ensemble.

Pressure Coupling

Analogous to the controlling of temperature detailed above, the pressure can, in a similar fashion, be controlled by coupling the system to a pressure bath. The pressure can be controlled by altering the box size of the system, with an expansion of the box reducing the pressure, and a contraction of the box increasing the pressure. The simplest procedure is the Berendsen algorithm,[68] where the coordinates and box vectors are rescaled at every step towards a given reference pressure, P according to

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_p} \tag{2.24}$$

where P is the current pressure, P_0 is the reference pressure, and τ_p is the time constant. A more advanced procedure, particularly useful for systems where fluctu-

ations in pressure are more important, is the Parrinello-Rahman barostat.[69] This approach works by updating the equations of motion for particles in every step.

2.4.6 Analysis techniques

Radial distribution functions

The radial distribution function (RDF) is a pair correlation function which describes the average structure of atoms in a molecular system. An RDF is constructed in a series of steps. Firstly, an atom (or group of atoms) are selected in the system. A series of concentric spheres, set at fixed distances (Δr) apart, are then drawn around the tagged atom (as shown in fig 2.7, below). Snapshots are taken of the system at regular intervals (typically for an equilibrated system) and a count done of the number of atoms in each shell. This data is then averaged to calculate the average number of atoms in each shell, and then divided by the volume of each shell and the average density of atoms in the system to give a radial distribution function. This can be mathematically expressed with the following equation

$$g(r) = \frac{n(r)}{\rho 4\pi r^2 \Delta r}, \tag{2.25}$$

where $g(r)$ is the RDF, $n(r)$ is the average number of atoms in a shell of width Δr at distance r, and ρ is the mean atom density.

A radial distribution function (with an example RDF shown in fig 2.7, below) typically has a $g(r)$ of zero at short separations, which indicates the effective width of the atoms as they cannot approach any closer. A number of obvious peaks appear which indicates the probability of finding an atom at that interatomic separation.

The structure of an RDF will vary significantly depending on whether the system is solid, liquid or gas. For a crystalline system (solids), the peaks are particularly sharp at both short and long range due to the strong confinement of atoms to their lattice positions.[70,71] For liquids, where disorder and continual movement of atoms is constant, an average structure is much more effective to work with.[72,73]

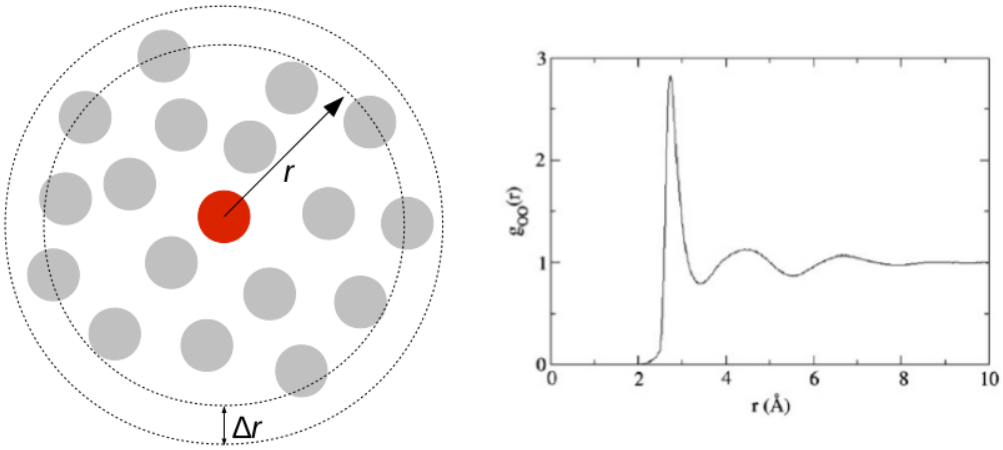

Figure 2.6: A snapshot of a system (left) with tagged molecule (red), and an example radial distribution function for oxygen-oxygen within water (right). RDF taken from Clark *et al.*[2]

Binding energy

The binding energy (or free energy of association) for two chromonic molecules can be evaluated from the potential of mean force (PMF) profile along the separation distance between the centres of mass (COMs). This was achieved using a set of constraint distance simulations. For each constraint distance between the two COMs, the molecule is left to find the lowest energy conformation, and the average constraint force is then evaluated. The PMF is then calculated by integrating the mean constraint force over the separation distance, s, using the following equation:[74–76]

$$V_{\mathrm{PMF}}(r) = \int_{r_{\max}}^{r_0} \left[\langle f_c \rangle_s + \frac{2k_B T}{s} \right] ds + C \tag{2.26}$$

where, $\langle f_c \rangle_s$ is the average constraint force between the two COMs, k_B is the Boltzmann constant, and T is the temperature. The integration limits are r_0, the smallest distance between molecules, and $r_{\max}$, the largest distance where the molecules are no longer interacting. The kinetic entropy term, $2k_BT/s$, accounts for the increased volume sampled at larger separations by the freely rotating molecule when the direc-

tion between the molecules is not fixed, and is therefore deducted from the average constraint force.

The kinetic entropy term can be derived as followed: For a constraint in a 3-dimensional system of two masses rotating at a fixed distance s, a thin spherical shell of volume $W = 4\pi s^2 \delta s$ is available. From Boltzmann's equation, the corresponding entropy[77] is

$$S = k_B \ln s^2 + C, \tag{2.27}$$

where C is a constant. The contribution to the free energy is

$$-TS = -2k_B T \ln s. \tag{2.28}$$

Hence, the entropic force is

$$F(s) = -\frac{d}{ds}(-TS) \tag{2.29}$$

$$F(s) = \frac{2k_B T}{s} \tag{2.30}$$

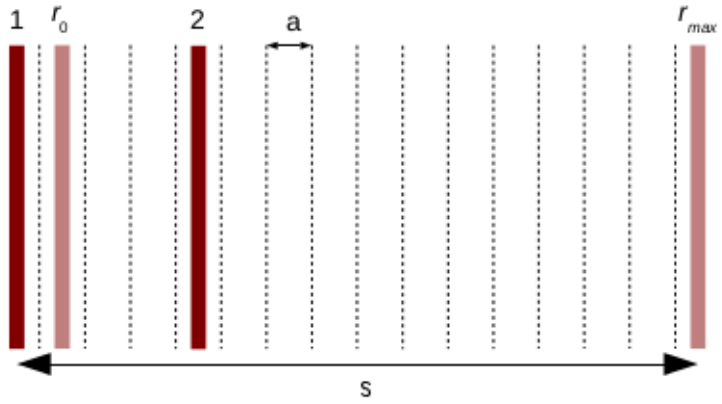

Figure 2.7: A schematic of the constraint distance simulations for a chromonic system, where (1) and (2) is an equilibrated dimer pair in solution, (a) is the difference between constraint differences, and (s) is the complete pull space sampled where r_0 is the minimum distance between molecules, and $r_{\max}$ is the maximum.

Cross sectional area

The cross sectional area (CSA) is the defined as the area of the cross section normal to the column axis. The CSA was calculated by determining the difference between the surface accessible area for a dye molecule within the stack (the surface accessible area for the sides of the dye molecule in the stack) and then as a molecule on the end of the stack exposed to the solvent. The CSAs were calculated by running a 0.14 nm spherical probe (correlating to the radius of a water molecule) over the selected structure using the molecular graphics program PyMOL.[78,79]

Stacking distance

The stacking distance , d, is defined as the inter-plane distance between neighbouring dye molecules within a column (how far are neighbouring dye molecules from each other in a stack):

$$d = \mathbf{d}_i \cdot r_{ij} \tag{2.31}$$

where r_{ij} is the vector between the COMs of molecules i and j, and d_i is the unit vector defining the normal to the plane of molecule i, defined by 3 selected atoms on the plane in a disc-shaped molecule. Neighbouring dye molecules were analysed within a range of 0.2-0.5 nm, whereas next-neighbouring dye molecules were analysed within a 0.5-0.8 nm range, for all dye pairs, $i \neq j$.

Twist angle

The twist angle, θ, refers to the angle between each dye molecule in a stack and its neighbouring molecule

$$\theta = \cos^{-1}(\mathbf{v}_i \cdot \mathbf{v}_j), \tag{2.32}$$

where $\mathbf{v}_i$ is the unit vector defined by

$$\mathbf{v}_i = \mathbf{L}_i \times \mathbf{d}_i, \tag{2.33}$$

where $\mathbf{d}_i$ is the unit vector defining the normal to the plane (discussed above for stacking distances), and $\mathbf{L}_i$, the unit direction vector defined by two atoms in the

plane of the disc-shaped molecule. Neighbouring dye molecules are analysed within a range of 0.2-0.5 nm, whereas next-neighbouring dye molecules were analysed within the 0.5-0.8 nm range, for all dye pairs, $i \neq j$.

Population of twisting angle along the stack

Pair correlation functions mapping the twisting angle and distance along the stack as a function of occurence was produced by incorporating equations (2.31) and (2.32) into a single histogram. All dye pairs were calculated for $i \neq j$.

Column-column distance

The column-column distance, I, refers to the distance between columns in a layer structure:

$$I = \mathbf{d}_i \cdot \mathbf{r}_{ij} \tag{2.34}$$

where $\mathbf{r}_{ij}$ is the vector between the COMs of molecules i and j. The unit vector defining the normal to the plane of molecule i, $\mathbf{d}_i$, is defined by 3 selected atoms on the molecule. All dye pairs were calculated for $i \neq j$.

Chapter 3

Atomistic studies of cyanine-based chromonic mesogens

3.1 Introduction

This chapter focuses on two chromonic mesogens: 5,5'-dimethoxy-bis-(3,3'-di-sulphopropyl)-thiacyanine triethylammonuim salt (Dye A) and 5,5'-dichloro-bis-(3,3'-di-sulphopropyl)-thiacyanine triethylammonuim salt (Dye B). The two mesogens share an identical molecular core, but differ in the peripheral groups attached to the aromatic rings (the 4-position of the external phenyl ring); Dye A has two methoxy groups, whilst Dye B has two chlorine atoms. Both dyes have the same counterion (triethyl ammonium cation). The molecular structure for both Dye A and Dye B is shown in fig 3.1.

Phase diagrams, constructed from x-ray diffraction and optical micrography experimental data,[3,4] for Dye A (fig 3.3) and Dye B (fig 3.4) shows the phase behaviour for each dye as a function of temperature and concentration of dye. The Dye A phase diagram shows the isotropic, nematic and smectic phases formed, as well as transitions between these phases. A few key features to note are that the nematic phase exists over a very limited temperature and concentration range, and that the smectic nucleates from the isotropic phase but not from the nematic. The major feature in the Dye B phase diagram is the extensive smectic phase, and the lack of any nematic phase.

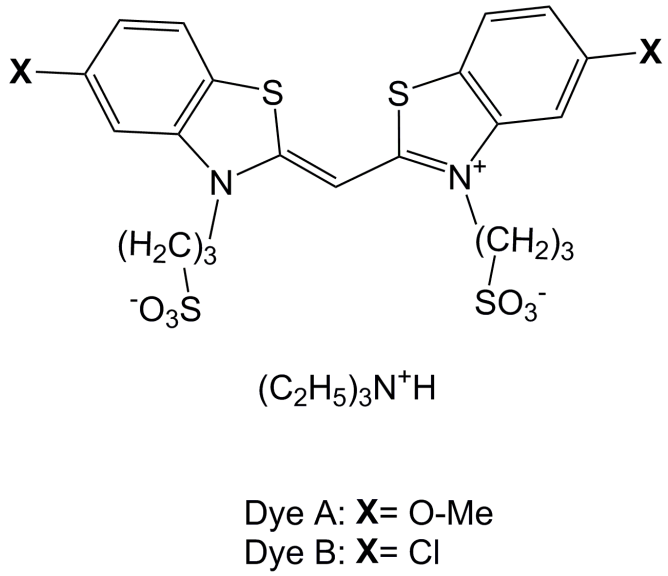

Figure 3.1: Molecular structure of Dye A and Dye B, and counterion triethyl ammonium. The central unit connecting the two larger aromatic regions in the dye is conjugated, making both dye molecules symmetrical.

The proposed structure for the nematic phase of Dye A differs from that of a standard J or H aggregate. X-ray studies[3,4] of Dye A have found a larger than expected aggregate cross section, which has led to the tentative assignment that this structure is a double-width column (fig 3.2). In the suggested structure, each 'layer' within the column is composed of two dye molecules, with the flexible arms pointing away from the center. The next layer is then rotated by 90° to avoid unfavorable interactions with the sulphonate groups of the neighboring layer. This structure has only been seen once before in the study of Bordeaux dye, a polyaromatic chromonic mesogen derived from napthalenecarboxylic acid.[5]

One of the main driving forces behind the selectivity for face-to-face aggregation is due to favourable interactions between aromatic cores. Quadrupolar interactions between the aromatic central units drive the formation of J and H aggregates. The molecular configuration of the molecules within the proposed model is not commensurate with the usual preferred quadrupolar interactions of the core, leading to speculation as to whether the experimental data could be re-interpreted. The similarity of the structures of Dye A and Dye B will allow for a direct study of how the molecular structure alters aggregation behaviour.

At higher concentrations for both Dye A and Dye B, a layer phase has been

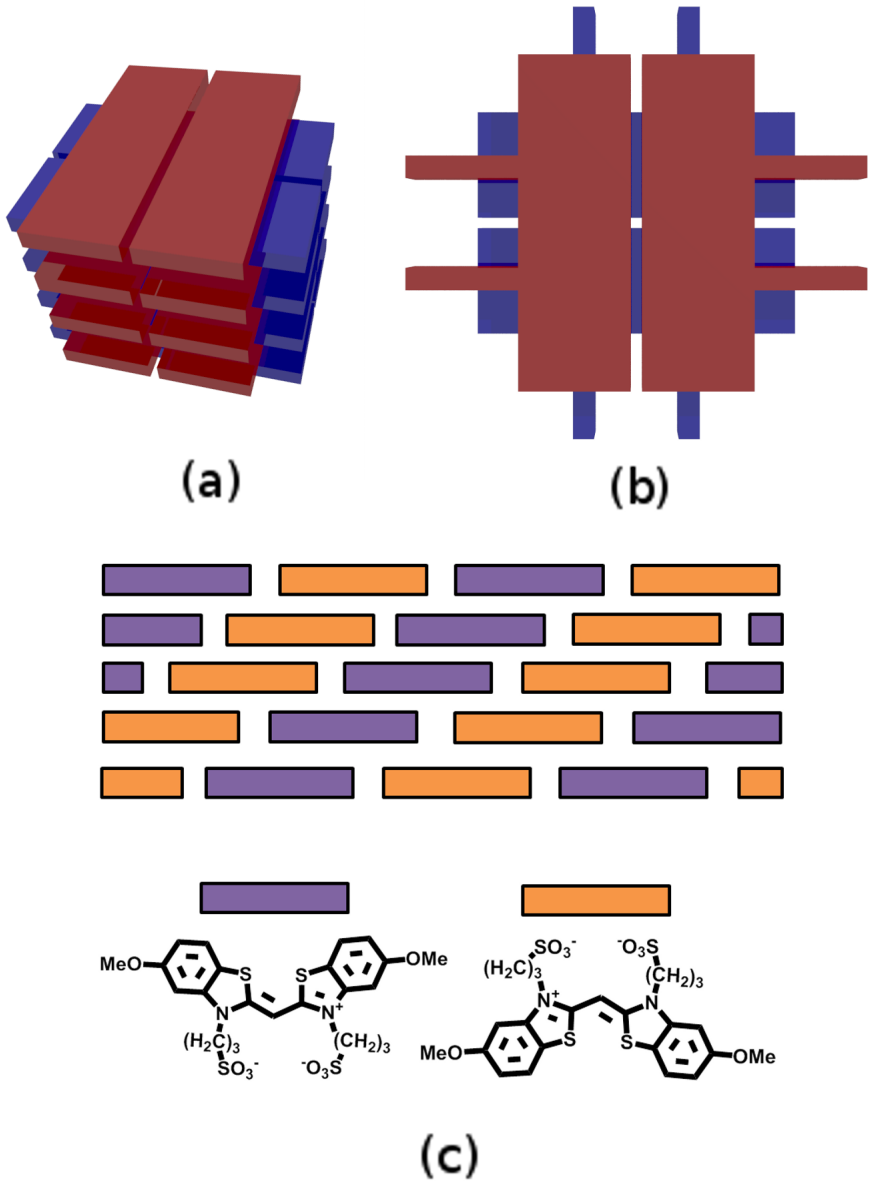

Figure 3.2: Proposed bimolecular stacking for Dye A in the nematic region, where the red and blue colour change indicates the next layer of the column. Side view (a) cross section of the column (b). Suggested brickwork packing arrangement within a single Dye A layer in the smectic region, whereby the dye molecules within the layer are stacked perpendicular to the plane of the paper.

proposed.[4,25] The proposed phase is built from adjacent molecules aligning in a brickwork fashion. The layer is also proposed to be one molecule thick. Both dyes are said to have a relatively high amount of positional order. There is long range order between layers, and short range order within the layer itself. Dye A is proposed to have a brickwork model, where neighboring dye molecules in a row of the layer have sulphonate groups alternating in or out of the plane of the layer. Long range order shows the system to be biaxial, with a repeat in layer structure every fifth row. Dye B is markedly different to the Dye A model, where each row within the layer alternates with either sulphonate groups pointing down or up out of the layer, as shown in fig 3.2. Contrary to the model for Dye A, Dye B is said to be uniaxial. A drawback is that the brickwork model struggles to accommodate a packing arrangement that does not have hugely unfavorable electrostatic interactions. This has led to speculation as to whether these non-columnar chromonic layers are feasible.

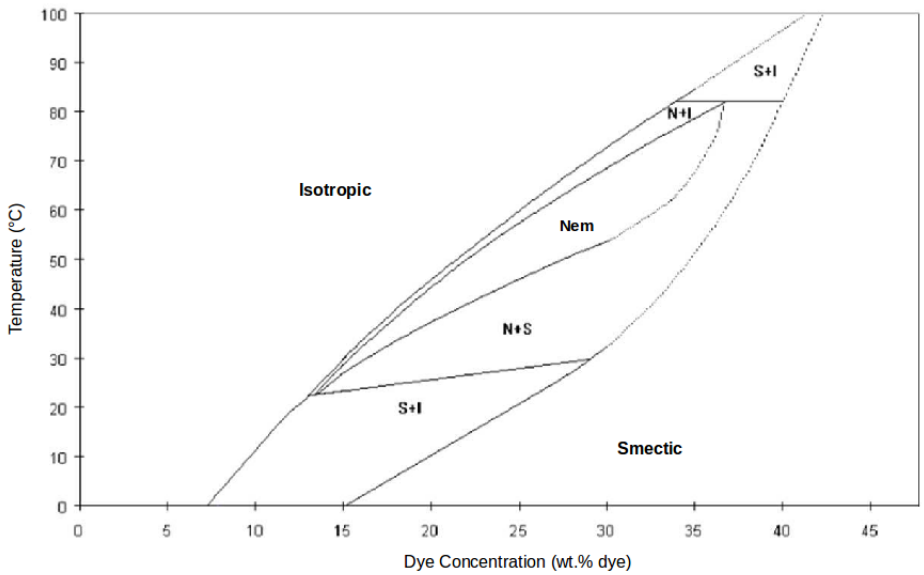

Figure 3.3: Phase diagram for Dye A. (Taken from Bottrill *et al.*)[3]

Molecular dynamics simulations can provide a molecular level picture of the preferred stacking arrangements and dynamics within an aggregate. A successful recent case study involves the computational study of the anionic azo dye sunset yellow.[47] Simulations were able to match key experimental findings, as well to provide insight into the preferred molecular orientation within aggregates. In this

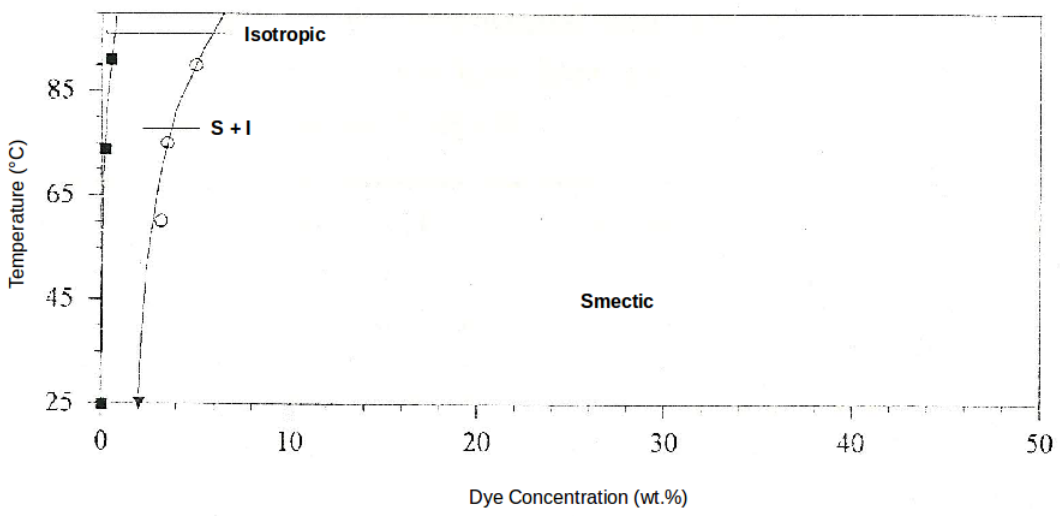

Figure 3.4: Phase diagram for Dye B. (Taken from Bottrill *et al.*)[4]

chapter we use a similar methodology to probe the self assembly behavior of Dye A and Dye B in aqueous solution, over a range of concentrations.

3.2 Computational details

The potential functions and interaction parameters were employed by the GAFF parameters[60,80] as implemented by the simulation suite Gromacs,[81–85]

$$E_{\text{total}} = \sum_{\text{bonds}} K_r(r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\theta - \gamma)] + \sum_{i<j} 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] + \sum_{ij} \frac{q_i q_j}{\varepsilon r_{ij}}. \tag{3.1}$$

The partial atomic charges were calculated from BCC partitioning with the AMBER sqm (semiemperical quantum mechanics) programme, which carries out single point calculations and energy minimisations (geometry optmisations). The torsional angle (ϕ) is non-standard due to the delocalisation across the central bridging carbons, hence a torsional energy profile was calculated using the Gaussian 03[86] and the hybrid density functional theory B3LYP[87,88] together with a 6-31G** basis set. As

demonstrated by the energy profile (fig 3.6), the presence of two wells at 0° and 120° show that the molecule has two low energy configurations, and all other deviations from these two angles are energetically unfavourable. The lowest energy configuration at 0° is a rigid, planar configuration, with the barrier height for rotation to the second configuration being approximately 70 $kJmol^{-1}$.

The GAFF parameters originally allowed full rotation around the torsional angle ϕ, and were therefore modified to the energy provided by the torsional energy profile (fig 3.6), hence ensuring a planar configuration of the core.

Molecular dynamics simulations were initially equilibrated in a canonical (NVT) ensemble and then followed by an extensive equilibration run using the isobaric-isothermal (NPT) ensemble. All simulations, unless otherwise stated, were performed at a temperature of 300 K, with a Nosé-Hoover thermostat[65] being employed. Pressure was controlled with a Parrinello-Rahman barostat[69] at atmospheric pressure with an isotropic pressure coupling. All simulations were initialized without bond constraints, until the simulation settled, bond constraints were then applied using the LINCS algorithm[89] and the time step increased from 1 fs to 2 fs. The interaction cut-off for Lennard-Jones interactions was 1.1 nm, while short range Coulombic interactions were truncated at 1.2 nm. The long range part of the electrostatic interactions was represented by employing a Particle Mesh Ewald (PME) summation.[55] The TIP3P[90] water model was chosen as most suitable for this system, due to accurate aqueous membrane bilayer simulations demonstrating that the GAFF/TIP3P combination could be utilized in biomolecular modelling.[91] Once equilibrated, a 100 ns production run was performed, and sampling was performed at 400 fs increments.

Initial simulations consisted of a system of 30 dye molecules and 30 counterion molecules randomly oriented and positioned in a cubic box. This process was followed by solvation of the system to give a 9 wt.% dye concentration. The same process was applied for concentrations of 20 and 30 wt.%. For the 30 wt.% solutions, additional larger simulations were performed consisting of 100 dye-counterion pairs in solution. For reference, any weight fraction concentration refers to the weight of both the dye and the counterion combined. A 9 wt.% concentration simulation

for both Dye A and Dye B showed the formation of two stacks with an antiparallel molecular arrangement. With diffusion of the two stacks in solution being slow, all evidence from simulations performed showed that a single column would form if allowed to run for a longer time period. It is for these reasons that for analytical purposes, a single seeded column of full anti-parallel arrangement will be used instead. Both of these seeded structures have been energy minimised using a steepest descent algorithm , and then allowed to anneal over many hundreds of nanoseconds, providing evidence for the stability of these anti-parallel columns. Seeded columns consisted of a system containing 30 molecules prearranged in an aggregate. Counterions and solvent were randomly distributed around the stack to make 9 wt.% solution. Columns were pre-assembled into both parallel and antiparallel configurations, where neighboring dye molecules are arranged in a head to head, or a head to tail arrangement respectively. For the antiparallel seeded columns, the initial interlayer distance was set at 0.35 nm, whereas for the parallel column, a greater interlayer distance was required to provide adequate separation between sulphonate groups. These distances were 0.55 nm and 0.4 nm for Dye A and Dye B respectively.

A double width column analogous to the model proposed by Bottrill[3] was also created by pre-assembling 32 Dye A molecules into a bimolecular stack. Each 'layer' within the column was composed of two dye molecules, with the flexible arms pointing away from the center. The next layer was then rotated by 90° to avoid unfavorable interactions with the sulphonate groups of the neighboring layer. The aggregate was solvated with the addition of TIP3P water and counterions around the column to create a stack at 26 wt.% concentration. The interlayer spacing was 0.4 nm. The larger distance was necessary to avoid overlap of sulphonate groups with methoxy groups in the adjacent layer. The temperature of the simulation was 326 K and pressure was atmospheric pressure, with the increase in temperature corresponding to conditions required for the nematic region of the phase diagram.

The free energy of association (using a PMF, as discussed in Chapter 2) was calculated for a 2 wt% Dye A dimer at 300 K. The Dye A system was equilibrated for 100 ns, before implementing "a pull" within the Gromacs package to calculate the PMF.[81] The initial configurations were generated with constrained distances

between the centre of masses of both molecules in the stack. A pull rate of 0.001 nm ps^{-1} was employed. For each chosen distance between centre of masses, the system was equilibrated for 1 ns, followed by a 50 ns production run where forces were outputted every 10 fs, for a simulation of 2.5 x 10^7 steps with a timestep of 2 fs. The distance between the centre of masses was varied from 0.2 to 2.0 nm, where spacing between the neighbouring points varied between 0.01 and 0.1, with smaller spacings for regions near the global minimum, and larger spacings as the dye molecules is pulled further out into solution.

Various analysis methods were used to quantitatively assess the structure of the aggregates. A detailed explanation of these various analytical techniques can be found in the end of Chapter 2. The distances between molecules in a stack are defined by the vector between the centre of mass of molecules. In order to have a stable centre of mass, only the core of the dye molecule was used in the analysis as shown in fig 3.5. The unit vector defining the normal to the plane of the molecule, was defined by atoms N, S and C2 (highlighted in red on figure 3.5). The unit direction vector (for the angle calculation) is defined by atoms C3 and C6 (highlighted in blue on figure 3.5). Neighbours were calculated within a 0.2-0.5 nm range, and next-neighbours within a 0.6 - 0.9 nm range. All distance and angle analyses was performed on equilibrated systems with a 10,000-20,000 frame trajectory. The cross sectional area (CSA) of the aggregates was calculated on the hydrophobic cores of the molecules only (thus matching the experimental assumption of the arms being fully solvated). A single frame from an equilibrated structure was chosen for each system being analysed.

3.3 Results

3.3.1 Short range order within the aggregates

Upon initial inspection the two dyes behave in an identical manner. Initial aggregation of the dye molecules occurs in under a nanosecond, from a fully dispersed system, dimers and trimers form quickly. The system evolves over several hundred nanoseconds to eventually form a large-single stack, as shown in fig 3.7, for Dye A.

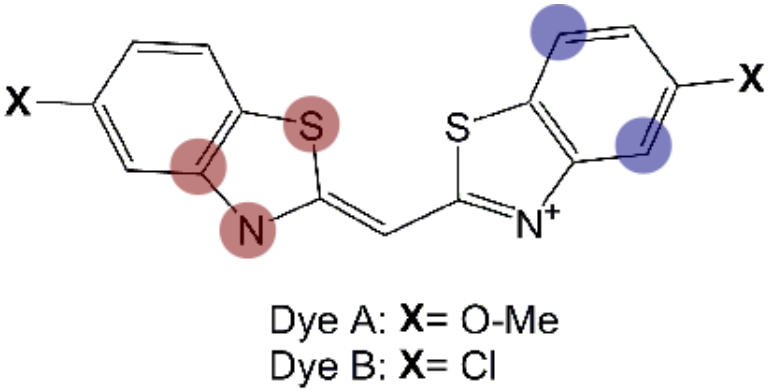

Figure 3.5: Molecular structure for the core of Dye A and Dye B, with the atoms used to define the plane of the molecule highlighted in red, and the atoms to describe the direction vector in blue.

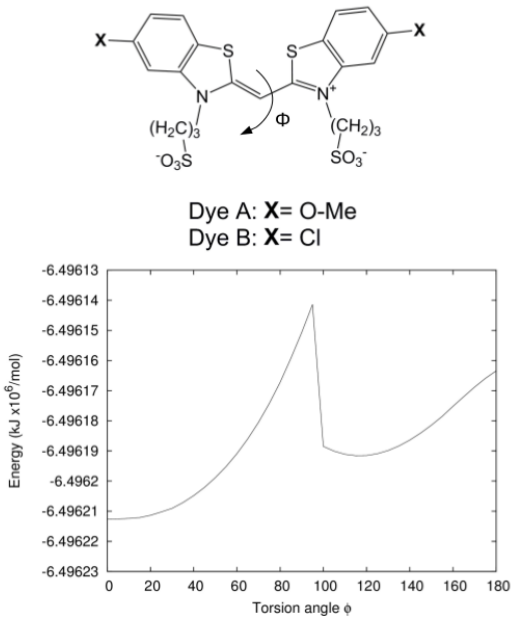

Figure 3.6: Energy as a function of rotation angle (bottom) around selected torsion (top).

All concentrations showed the same strong affinity for self-assembly, with no isolated monomers/dimers/trimers present in the system at equilibrium. Similarly, at all concentrations there is a strong preference for local order within a single column; a predominant head-to-tail packing arrangement. All concentrations show the formation of a single width column. No intercalation is observed between smaller aggregates, neither is there any back-to-back stacking between dye molecules. Face-to-face aggregation is observed as the predominant interaction between dye molecules.

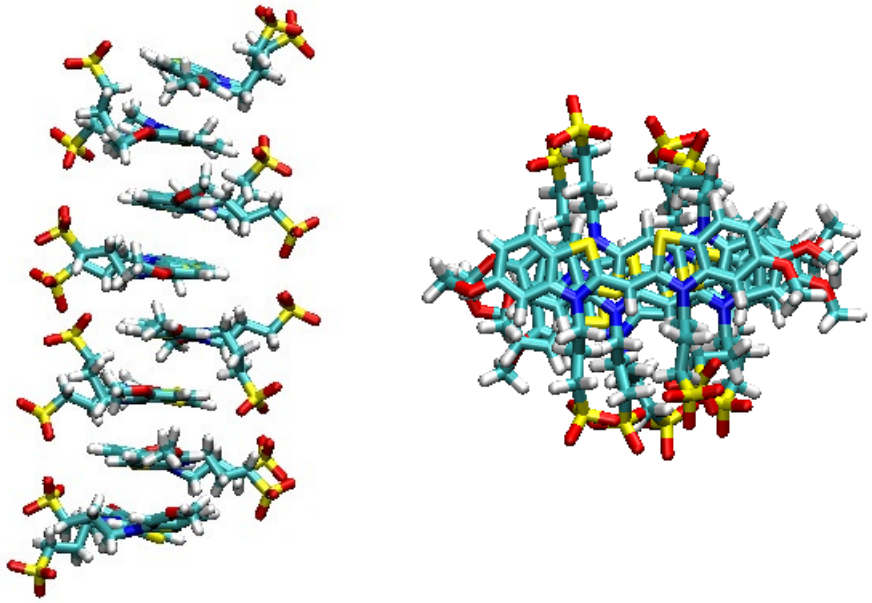

Figure 3.7: Short range order for the Dye A aggregate. Side view showing antiparallel arrangement between dye molecules (left) and top view showing the cross section of the column (right). Solvent and counterions are not shown. This short range order is seen for both Dye A and Dye B.

Aggregation between dye molecules shows a strong preference for an anti-parallel configuration, where the sulphonate chains alternate in direction through the stack, thus minimizing electrostatic repulsion that would otherwise occur. A single aggregate can be visualized as a column with an aromatic hydrophobic core, with two densities of ionic groups pointing into the polar solvent. Despite a strong configurational preference for antiparallel stacking, the drive to self-assemble can outweigh the energy penalty associated with a parallel configuration. The molecular strain in an aggregate caused by parallel stacking can be relieved by one of two possible mechanisms (illustrated schematically in fig 3.8).

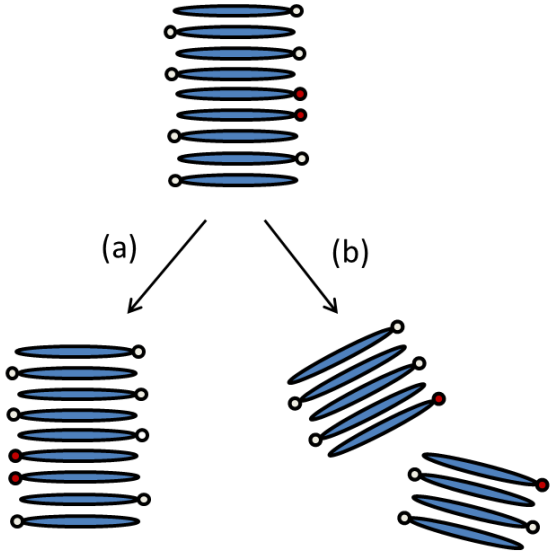

Figure 3.8: Routes available to eliminate parallel stacking between neighbouring dye molecules within an aggregate. (a) Single molecule rotation, generating another parallel configuration as a result. (b) Fragmentation of column, followed by eventual recombination to form an antiparallel arrangement.

One mechanism is where one of the dyes in a parallel configuration travels towards the end of the column where it can be eliminated. The other requires a large rotation of the sections of the aggregate either side of this molecule. It is mechanism (b) which has been observed for both dye systems. Parallel dye pairs appending a column are far easier to re-orient, with the dye molecule having a far greater degree of rotational freedom than a trapped molecule in the center. Substantial continuous parallel stacking of 4-5 adjacent dye molecules has also been observed and leads to a localized bend in the column. A strained-bent aggregate ultimately fragments and produces two smaller aggregates. Diffusion of the aggregates within solution eventually leads to recombination, in an attempt to form a fully anti-parallel stacking arrangement between neighboring dye molecules in a column. An example of this process can be seen in figure 3.9.

3.3.2 Free energy of binding

The free energy of association for a molecule in an aggregate relative to a molecule in solution can be seen from the PMF curve for a Dye A dimer at 300 K in fig 3.10.

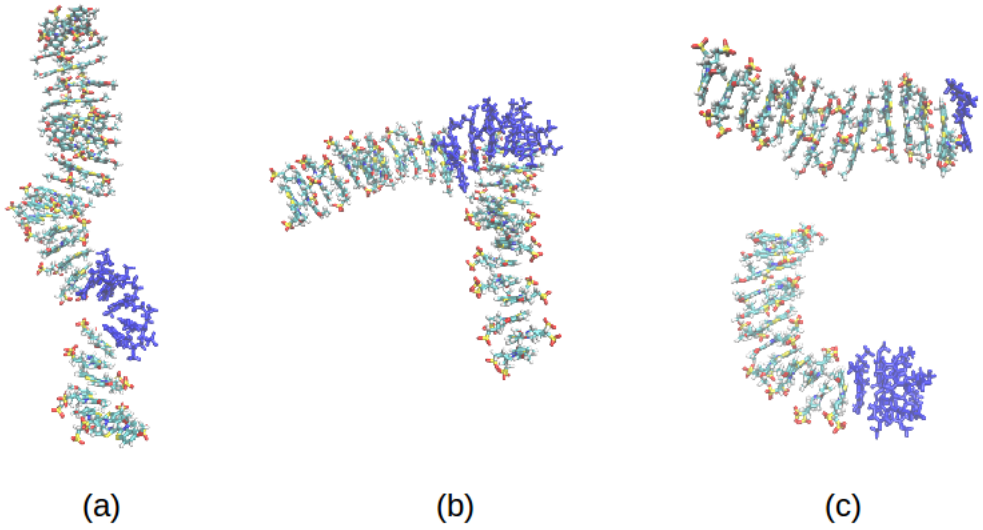

Figure 3.9: A demonstration of the fragmentation process for a 30 molecule stack for Dye A 20 wt.%. The highlighted region (blue) shows 5 molecules stacked in an parallel configuration, with fragmentation the chosen method of relieving the molecular strain of this packing.

The binding free energy for Dye A is estimated to be 14 k_BT (34.72 kJmol^{-1}). Dye A has a strong binding affinity relative to the typical binding energies (7-14 k_BT) reported in the literature for other chromonic molecules, such as TP6EO2M (12-14 k_BT,[52,53] Sunset Yellow (7 k_BT),[47] Disodium cromoglycate (7 k_BT),[92] Benzopurin (10.2 k_BT)[93] and Blue 27 (12 k_BT).[92]

3.3.3 Long range order within aggregates

The long range order within the aggregates, arising from the two different dye molecules, differs significantly. Dye A displays a clear chiral twist at the 9 and 20 wt.% solutions (fig 3.11). It should be noted that the ends of the aggregate do not meet through periodic boundary conditions to impose chirality due the box size. Despite a strictly achiral molecule and environment, spontaneous symmetry breaking occurs resulting in a chiral aggregate. 30 wt.% Dye A solutions show a non-chiral aggregate (fig 3.11).

Visual analysis of the 30 wt.% solution shows that the column interacts with adjacent columns via periodic images to form a layer structure (fig 3.12). The

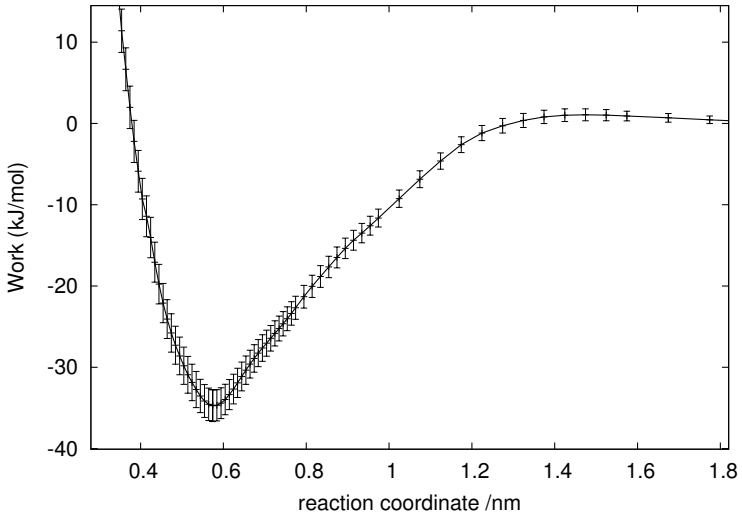

Figure 3.10: PMF curve for association for a Dye A dimer in solution.

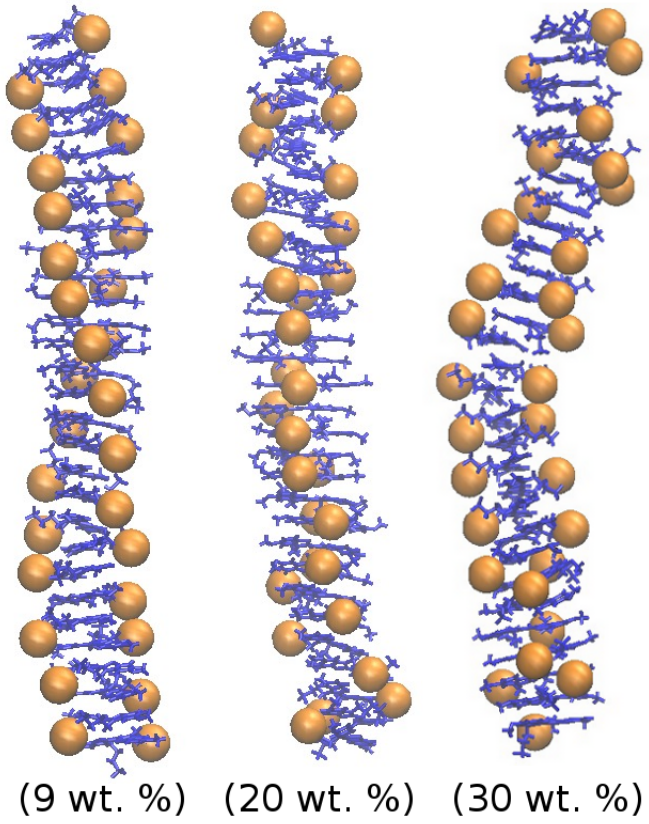

Figure 3.11: Chiral structures formed by Dye A aggregates at low concentrations (9 and 20 wt.%), and loss of chirality at a higher concentration of 30 wt.% A sulphur atom from the sulphonate group is highlighted in orange to make the twisting nature more visible.

transition from a chiral to a non-chiral column allows the highly charged sulphonate groups to minimize interaction with neighboring columns and the drive to maximize exposure of the sulphonate groups to the polar solvent environment, as opposed to being trapped in the apolar center of neighboring columns. The driving force behind the formation of the chiral aggregate for this dye system must be that a twisted state is the most energetically favorable conformation for non-interacting columns in solution. Chirality is absent for any concentrations of Dye B systems studied, as visually shown in fig 3.13. Higher concentrations of Dye B at 30 wt.% shows the formation of not one single column, but many smaller columns interacting to form a sheet-like structure.

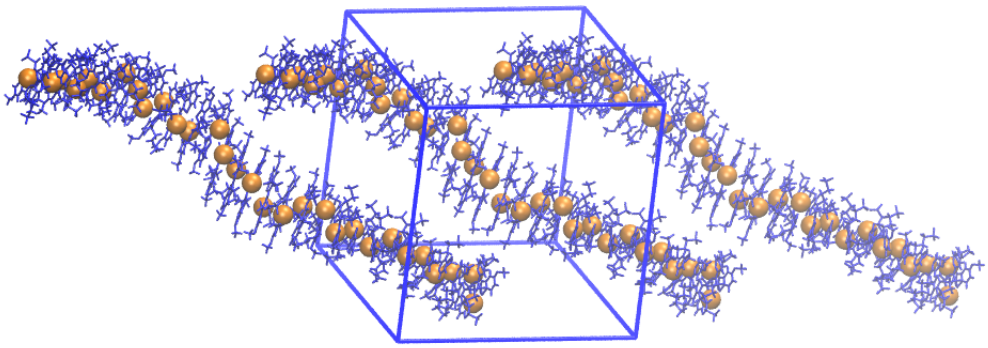

Figure 3.12: Achiral structure of Dye A at 30 wt.%, with both the periodic box shown in blue, and the periodic images of the stack in both directions of the Z-axis.

Larger systems containing 100 dye molecules for both Dye A, and Dye B, at 30 wt. % solutions, both show layer structures, matching the non-chiral structures seen in the smaller 30 dye molecule systems. Dye A showed the formation of 3 non-chiral columns, with the side to side interactions via the methoxy groups of these aggregates forming a layer structure in a plane. Dye B shows a network of smaller aggregates interacting via side-to-side and face-to-side interactions, forming a layer structure (both to be discussed later).

Interlayer distance analysis

The interlayer distance is the average distance between neighbouring dye molecules in a stack. Analysis of both Dye A (fig 3.14) and Dye B (fig 3.15) aggregates show

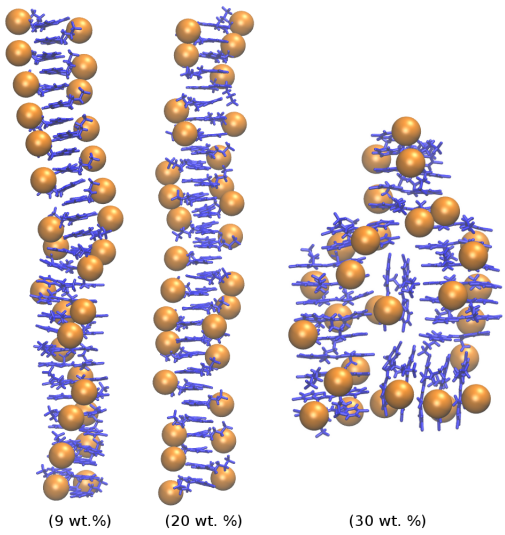

Figure 3.13: Achiral structures formed by Dye B aggregates at all concentrations simulated (9, 20 and 30 wt.%).

the average interlayer distance to be 0.35 nm, independent of the concentration. X-ray studies of chromonic dye systems show a typical peak corresponding to the 0.34 nm repeat distance, thus closely matching the interlayer distance found for the Dye A and Dye B systems.

Analysis of twist angle

Analysis of Dye A, calculating a histogram of population as a function of twisting angle between adjacent and next-adjacent molecules was used to observe how much molecules twist with respect to neighbours, and next neighbours (fig 3.16).

For all concentrations, neighbouring molecules within a stack show a preferred twist angle of 156-159°, corresponding to neighbouring pairs aligned in an antiparallel configuration. 30 wt.% shows a further smaller region of twist at a twist of 15-40°, resulting from the fact that the aggregates formed at this concentration have regions of parallel stacking. Analysis of next- neighbour molecules within a stack shows a preferred twist that varies as a function of concentration. 9 wt.% shows a twist of 11-17°, 20 wt.% shows 15-20° twist, and 30 wt.% shows a twist of 18-27°. All concentrations, despite having marginally different twist values, correspond to a parallel configuration between next neighbour molecules. 30 wt.% shows a further smaller region of twist from 130-170°, which as previously discussed, is due

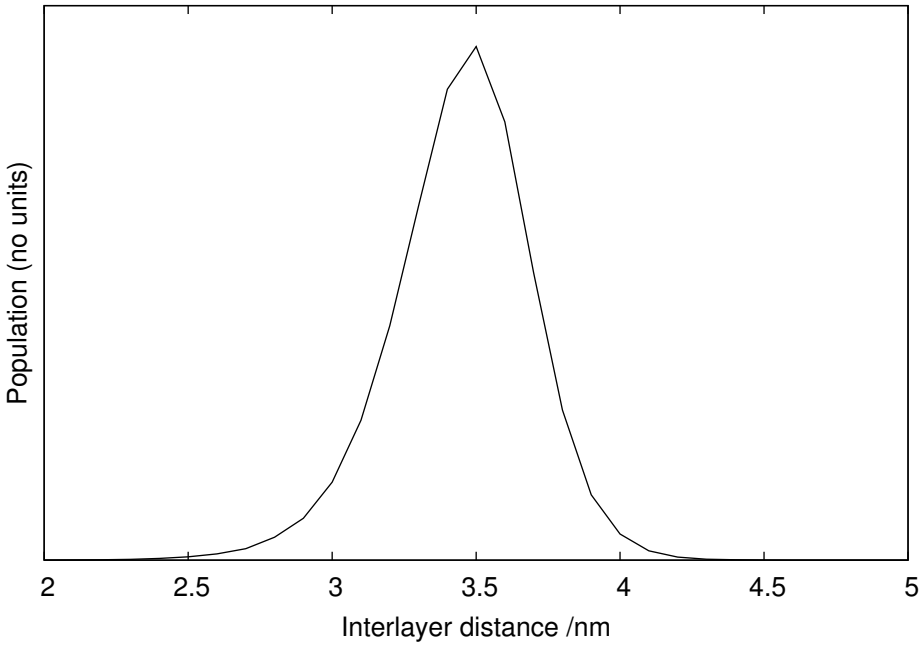

Figure 3.14: Interlayer distance for neighbouring Dye A molecules in a stack at 9 wt.% (black). The same stacking behaviour is observed for both the 20 wt.% and 30 wt.% Dye A systems.

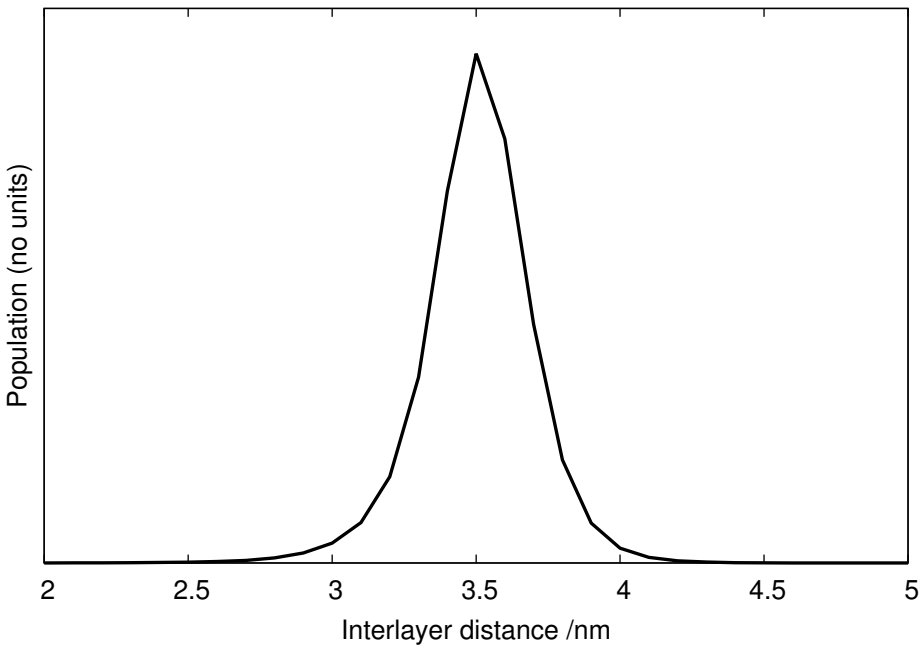

Figure 3.15: Interlayer distance for neighbouring Dye B molecules in a stack at 9 wt.% (black). The same stacking behaviour is observed for both the 20 wt.% and 30 wt.% Dye B systems.

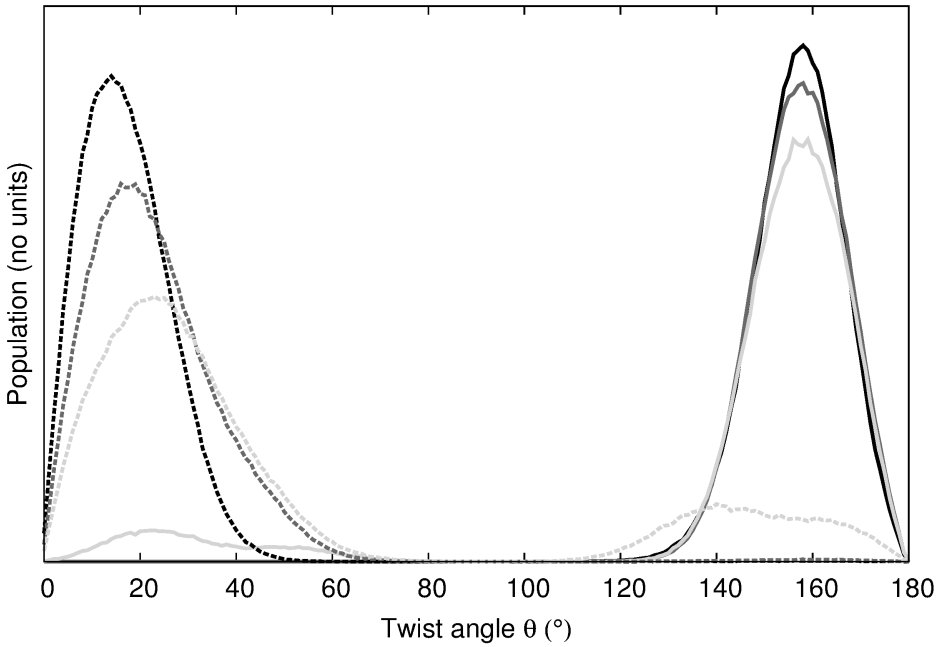

Figure 3.16: Histogram of twisting angle populations for Dye A aggregates for neighbouring dyes (solid), and next neighbours (dashed). 9 wt.% (Black), 20 wt.% (Grey), 30 wt.% (light grey). The difference in peak heights for neighbouring dye molecules is attributed to difference in the number of frames used to analyse each system.

to this concentration having regions of parallel packing between dye molecules. A large region of the histogram, from 70-110°, shows the absence of any twist for neighboring dye molecules, indicating that the shape of the central aromatic core strives for maximum overlap, restricting configurations to either parallel or antiparallel arrangements. The presence of the charged sulphonate moeties biases the orientation of the dyes within a stack to form with a predominantly head to tail arrangement. The effect of concentration on the average twist angle for Dye A appears to have little effect. The preferred twist angle shows minimal fluctuation as a function of concentration.

An analysis of Dye B twisting angle at 9, 20 and 30 wt.% has also been carried out (Fig 3.17).

Neighbouring dye molecules for all concentrations show a preferred twist angle of 169-173° (corresponding to an antiparallel configuration), while next neighbours show a preferred 6-12° twist (corresponding to a parallel configuration). Neighbour analysis for 9 and 30 wt.% show a small peak at 14-22° and 8-15° respectively, indi-

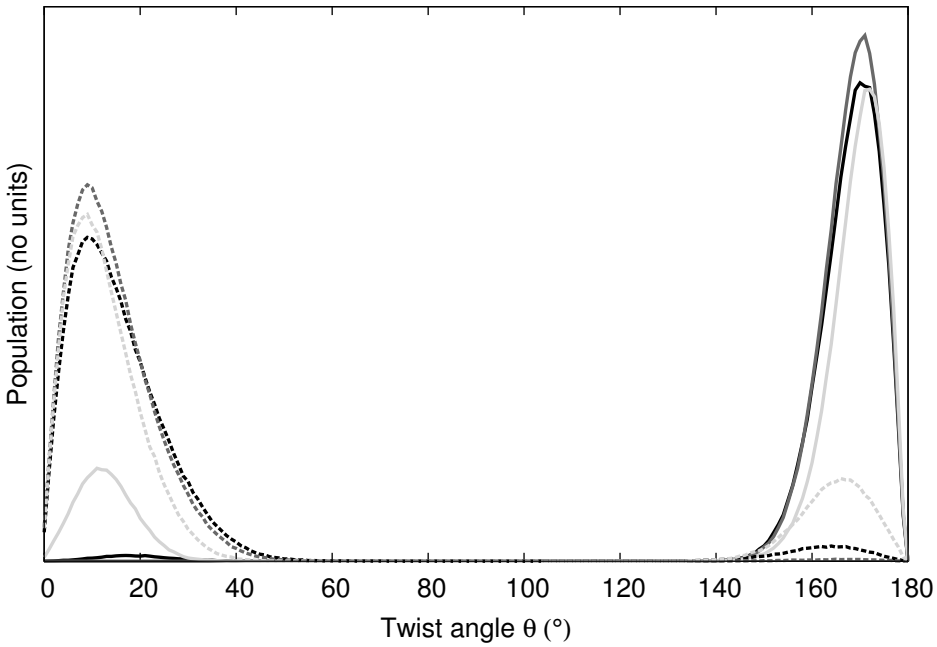

Figure 3.17: Histogram of twisting angle populations for Dye B aggregates for neighbouring dyes (solid), and next neighbours (dashed). 9 wt.% (Black), 20 wt.% (Grey), 30 wt.% (light grey). The difference in peak heights for neighbouring dye molecules is attributed to difference in the number of frames used to analyse each system.

cating a small degree of parallel configurations between neighbouring dye molecules within a stack. A small peak at 155-170° and 162-170° can also be seen for these concentrations in next-neighbour analysis.

Comparison of Dye A to Dye B twist angles yields two key differences. Comparison of neighbors as an example shows that Dye A, for all concentrations, has a larger offset twist angle of 156-159°, whilst Dye B has a twist angle of 169-173°. This arises from the bulkier methoxy substituents on Dye A, in comparison to the chlorine substituents on Dye B. The other key difference is the distribution of the peaks themselves, with Dye A having extremely large distributions, whilst Dye B has a sharper smaller distribution of angles.

Long-range chirality analysis

To analyze the long range order of Dye A, pair correlation functions mapping the angle and distance as a function of occurrence were calculated (Fig 3.18). This allowed for the helical nature of the columns to be analyzed. All concentrations

show that the peaks at each packing distance (increments of 0.36 nm) have a single distribution. Parallel arrangements from next neighbours are observed at distances of multiples of 0.72 nm, whereas antiparallel stacking for neighbours is found at distance of multiples of 0.36 nm.

9 and 20 wt.% solutions both show the intersection of both densities as a function of stack distance, thus confirming a helical column. The intersection point, calculated by interpolation of the graph, is approximately 4.5 nm, taking into account head-to-tail symmetry. This value corresponds to the $\frac{1}{2}$ pitch length of the helix. The 30 wt.% solution is markedly different from the other concentrations. No helical twist is observed, as noted by the two densities failing to intersect as a function of distance along the stack.

To also analyze the long range order of Dye B, pair correlation functions mapping the angle and distance as a function of occurrence were calculated (Fig 3.19). Chirality was absent in both high and low concentration systems of Dye B.

The difference in the preferred stacking arrangements for both dyes, i.e. the formation or absence of a chiral columnar aggregate, can be directly attributed to the change in molecular structure between Dye A and Dye B. The presence of the periphery methoxy groups must be crucial in causing Dye A molecules to stack at a persistent twist angle with respect to neighbouring dye molecules, due to a combination of the steric effect of the methoxy group between neighboring dye molecules, or the electronic properties of the aromatic core being affected by the presence of the methoxy group.

Cross sectional area analysis

The cross sectional area (CSA) for all aggregates in all the systems were calculated. No experimental CSA's have been provided in the literature for Dye B. All values can be seen in table 3.1. It should be noted that following the experimental assumption, CSA calculations for simulation results have been performed using the hydrophobic core of the dye only, eliminating the flexible sulphonate arms.

Dye A shows a concentration-independent CSA for all aggregates, given the relatively small fluctuation between values. There is no difference in CSA when

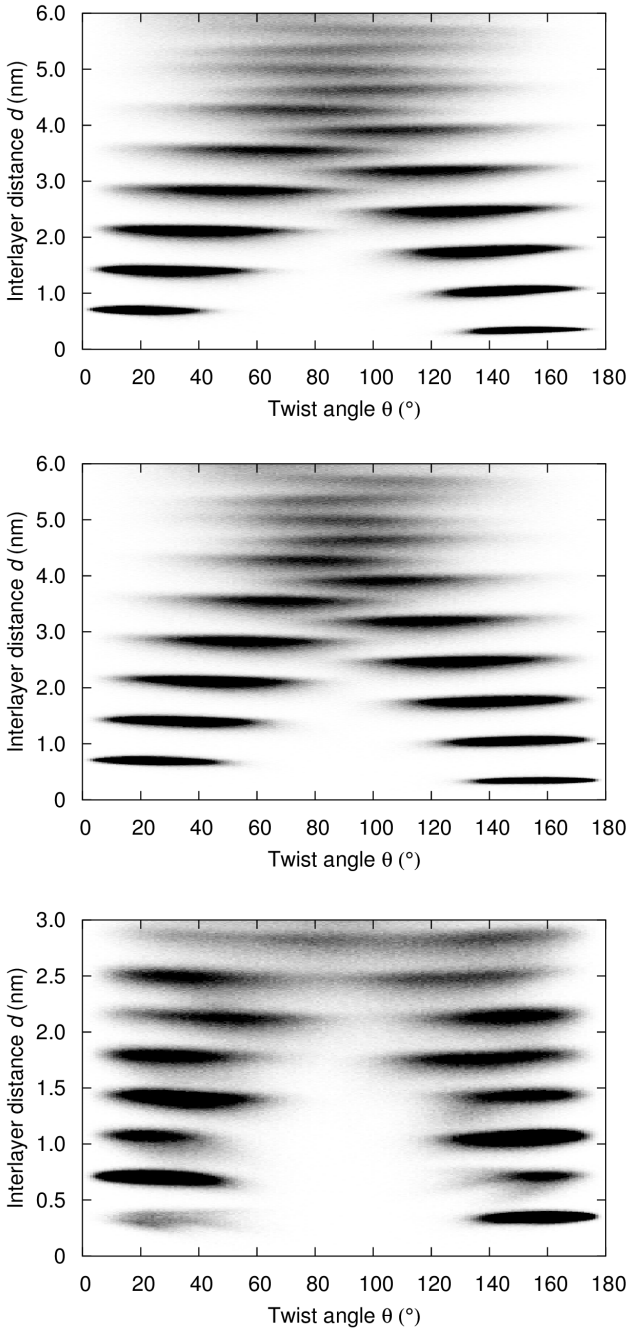

Figure 3.18: Histogram of twisting angle between all pairs and distance along the stack as a function of probability for Dye A at 9 wt.% (top), 20 wt.% (middle), 30 wt.% (bottom).

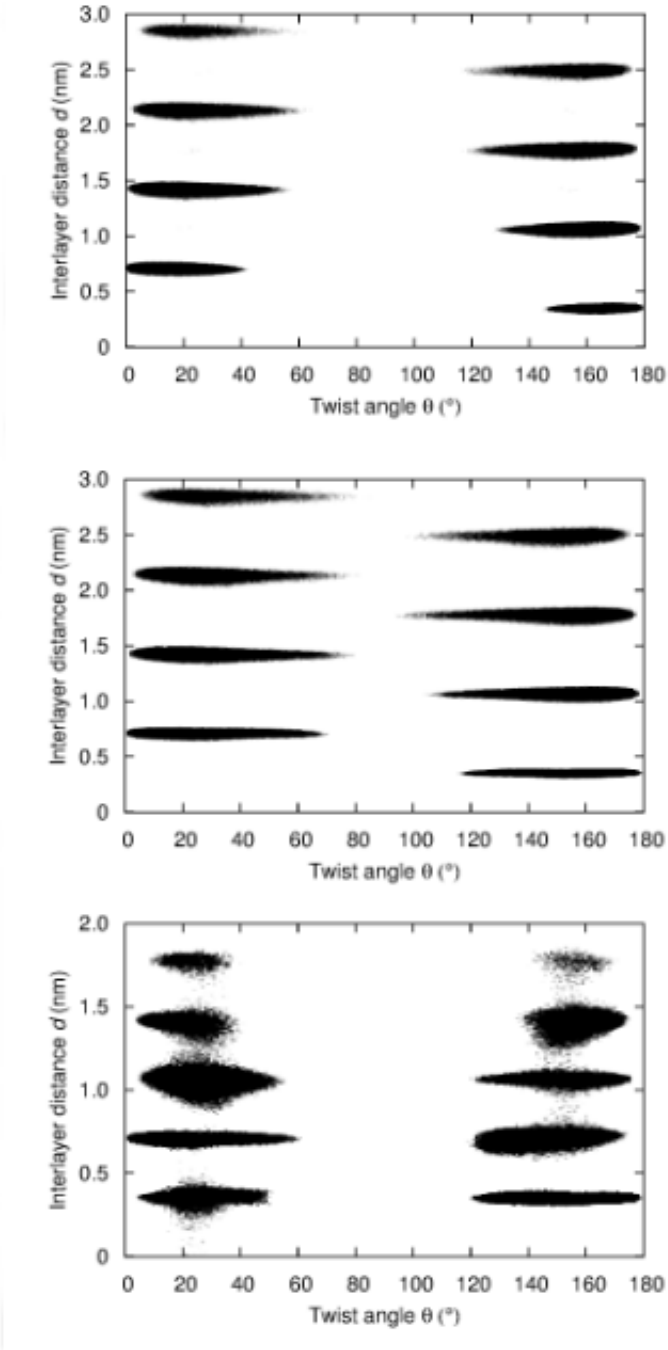

Figure 3.19: Histogram of twisting angle between all pairs and distance along the stack as a function of probability for Dye B at 9 wt.% (top), 20 wt.% (middle), 30 wt.% (bottom).

moving from a twisted structure (9 and 20 wt.%), to a non-twisted structure (30 wt.%). A comparison of Dye A and Dye B CSAs shows a slightly smaller value for Dye B aggregates. This can be attributed to the difference in the periphery groups of the core, where Dye A has a methoxy unit on either side, and Dye B has chlorine atoms.

Dye A results from simulation show a successful match to experimental values. The experimental aggregate structure suggested, a double width column, was based heavily on the experimental CSA values found. Simulation results have shown that a single width column, both twisted and untwisted, matches the experimental values, thus providing a reinterpretation of the aggregate structure for Dye A.

System	Dye A /$\mathring{A}^2$	Dye B /$\mathring{A}^2$
Exp (15 wt.%)	160	-
Exp (20 wt.%)	148	-
Exp (30 wt.%)	177	-
Sim (9 wt.%)	160.3 ± 5.7	141.0 ± 7.2
Sim (20 wt.%)	166.8 ± 1.4	137.2 ± 6.8-
Sim (30 wt.%)	161.7 ± 2.4	-

Table 3.1: Cross sectional areas for selected experimental concentrations (Exp) from the literature and simulation results for Dye A and Dye B at all concentrations.

Double-width Column Analysis

It should be noted that throughout the simulation trajectory, there is no evidence for the stability of a double width column. Both Dye A and Dye B show only face to face aggregation, resulting exclusively in columns with only one molecule per 'layer'. To further probe the proposed double width model for Dye A, a seeded double width structure was simulated. The seeded structure was formed by placing two dye molecules in a single 'layer', with sulphonate groups pointing away from the center. The next 'layer' above is then rotated by 90° to form the proposed model suggested in the literature. This structure started to fragment in under a nanosecond, eventually resulting in two separate single width columns, as observed for the unseeded simulations (fig 3.20). The breakdown of an idealised double-width column structure gives further credence to the hypothesis that this structure is not

energetically favourable, and the molecules prefer to adopt a single-width column (one dye molecule per row).

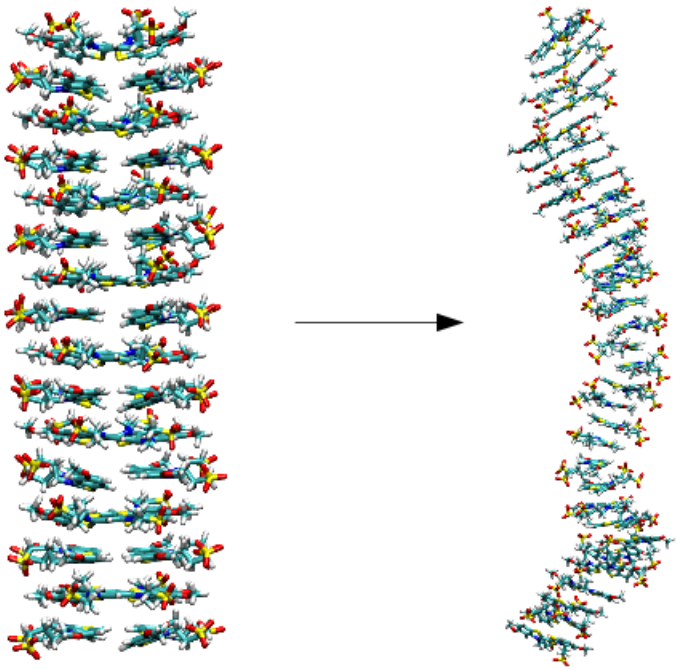

Figure 3.20: Initial seeded double width column structure (left) based on the suggested structure from Bottrill[3] and single width column formed after allowing the system to evolve over time (right).

3.3.4 Layer phase structure and formation

The alignment of columns in a plane to form a one-molecule thick layer structure is observed for both dye systems at higher concentrations (30 wt.%, 100 dye molecule system). The layer can be viewed as hydrophobic cores in the centre of the layer, shielded from the solvent by the position of the sulphonate arms on either side. The only difference is that the orderings of the columns within the layer is biaxial for Dye A (fig 3.22), and uniaxial for Dye B (fig 3.23). This is seen both from experiments and simulation.

The process of forming a layer, from a fully dispersed starting configuration, can be seen in a series of snapshots for the Dye B 30 wt.% 30 dye molecule system

(fig 3.21). The dye molecules aggregate to form one column in solution, as seen in 3.21 (a). Shorter columns then break off from unstable regions of the main column, and reinteract with the main column somewhere else 3.21(b). This process then continues until a layer structure is formed 3.21(c), which them retains this structure for many hundreds of nanoseconds. In this structure, sulphonate groups either point below or above the plane of the layer. The process of forming one or many larger aggregates, which then fragment and then interact with other aggregates to form layers, is seen for both Dye A and Dye B.

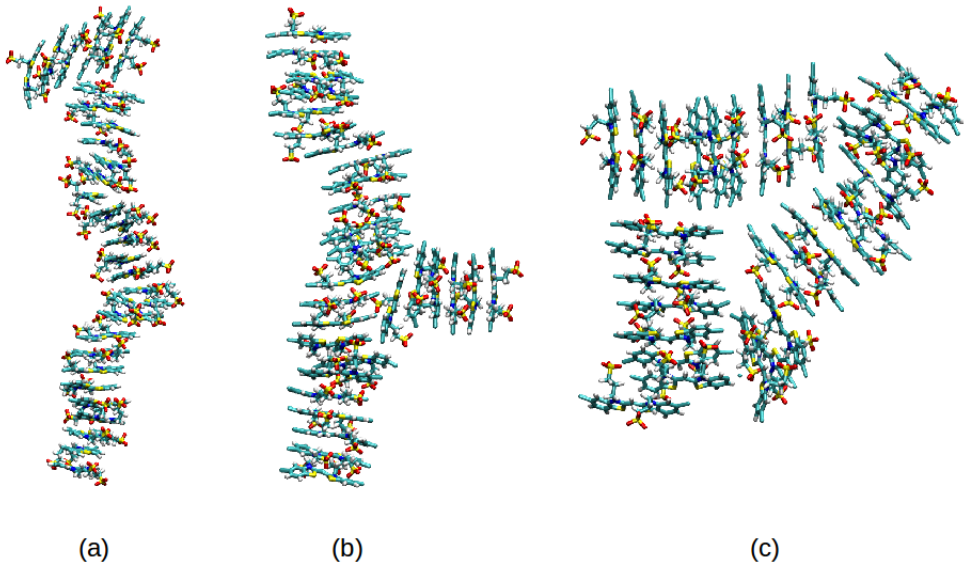

Figure 3.21: The formation of a layer structure for Dye B 30 wt.% 30 dye molecule system. A series of snapshots have be selected at various intervals: (a) 40 ns, (b) 80 ns, (c) 120 ns. Counterions and water have been removed for visual clarity.

Interlayer distance analysis

Interlayer spacing between chromonic dyes within an aggregate, in both Dye A (fig 3.24) and Dye B (fig 3.25) layer structures, have an average distance of 0.35 nm. This is in perfect agreement with the WAXS (Wide angle scattering) measurements made experimentally, where a 0.35 nm reflection was established to be present.[4]

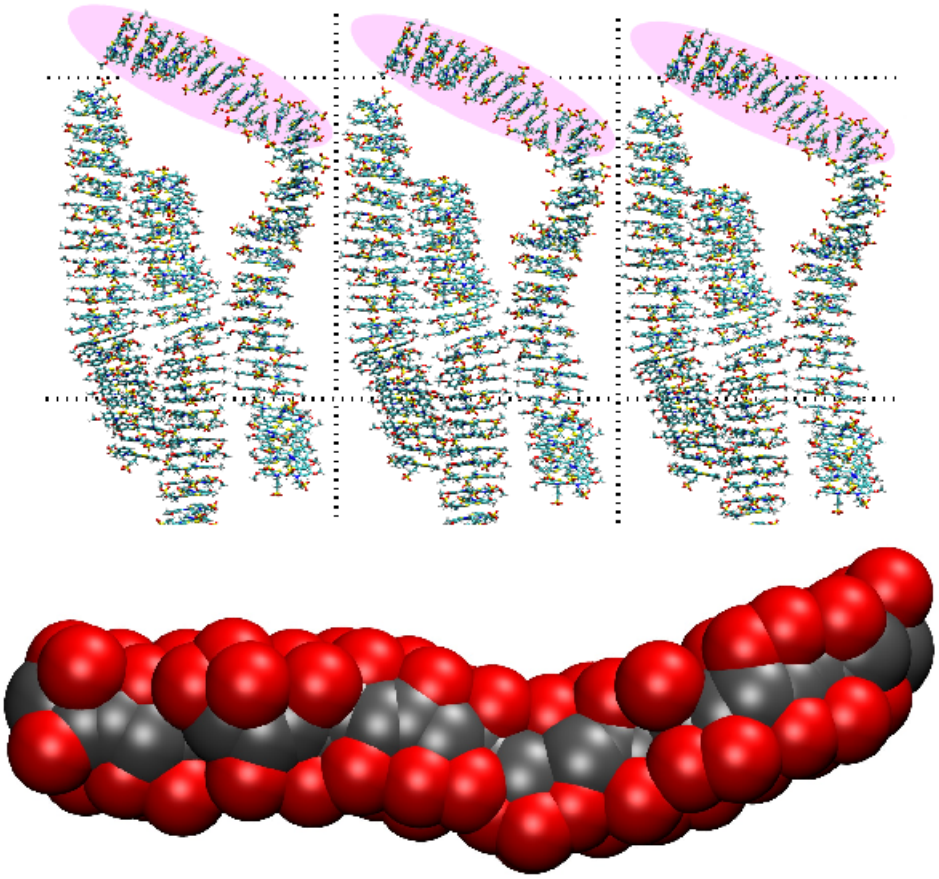

Figure 3.22: Layer structure for Dye A at 30 wt.% for 100 dye–counterion pairs in solution. Periodic boundary conditions are active to show the extent of the layer structure. The highlighted pink region indicates a region of an aggregate which curves out of the plane of the layer. The bottom structure shows the shielding nature of the sulphonate arms (red), pointing out into solution, and minimising exposure of the hydrophobic cores (grey).

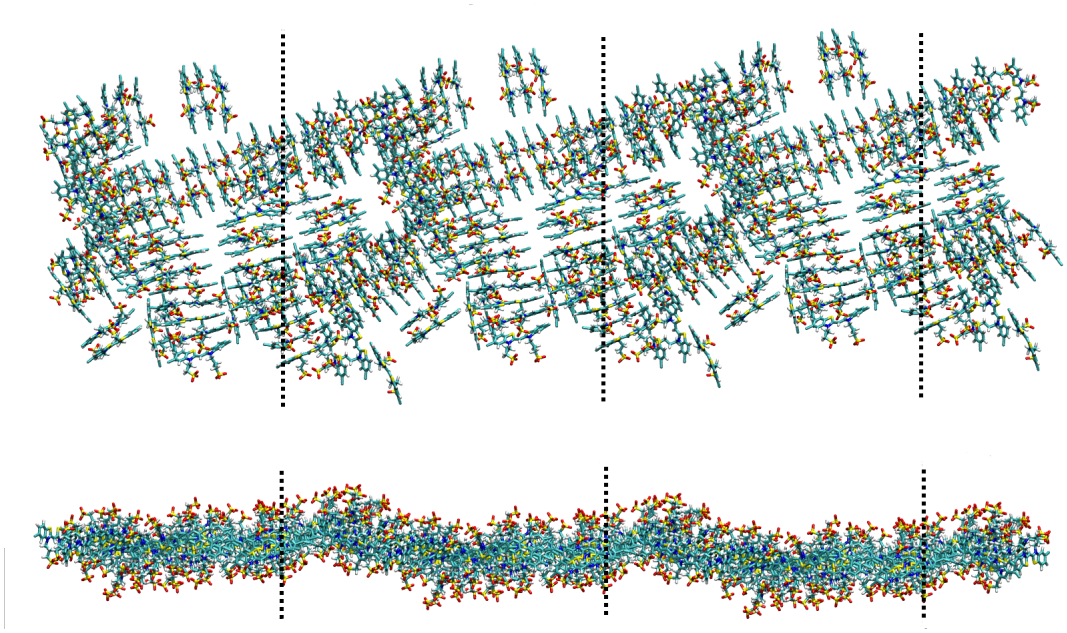

Figure 3.23: Layer structure for Dye B at 30 wt.% for 100 dye molecules in solution. Periodic boundary conditions are active to show the extent of the layer structure. A sky view (top) shows how the order of the columns within the layer is isotropic. A side view (bottom) shows how the layer is one molecule thick, and how the hydrophobic core is sandwiched between regions of sulphonate arms.

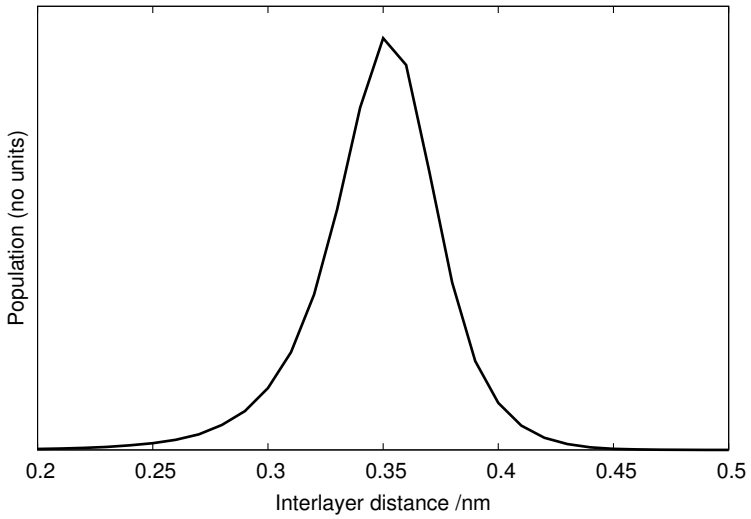

Figure 3.24: Interlayer distance between neighbouring Dye A molecules in a stack within the layer structure.

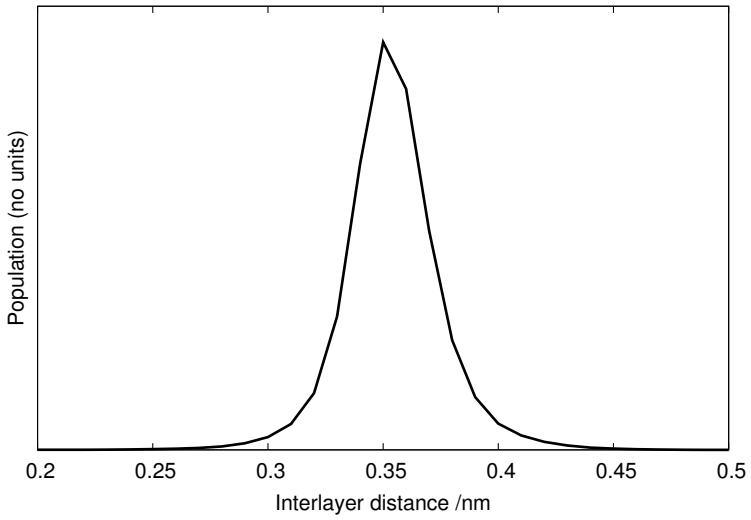

Figure 3.25: Interlayer distance between neighbouring Dye B molecules in a stack within the layer structure.

Analysis of twist angle

The neighbouring molecules within the stacks of the Dye A layer structure show a preferred twist angle of 155-162°, corresponding to an antiparallel configuration, in agreement with the twist angle observed in smaller systems of this dye. Next-neighbours show a preferred twist angle of 14-22°, corresponding to a parallel configuration between next-neighbours, and again agree with results found for smaller non-layer systems (fig 3.26).

The neighbouring molecules within the stacks of the Dye B layer structure show a preferred twist angle of 166-174°, corresponding to an antiparallel configuration, and next-neighbours showing a twist angle of 6-12°, corresponding to a parallel configuration between next-neighbours (as shown in fig 3.27).

Both Dye A and Dye B retain the average twisting behaviour between neighbours and next-neighbours when in an isolated stack, and when several stacks are interacting to form a layer.

Analysis of long-range chirality

Pair correlation functions mapping the angle and distance as a function of occurrence were calculated for Dye A (fig 3.28) and Dye B (fig 3.29) layer systems. In contrast to lower concentrations of Dye A, both systems show the formations of aggregates

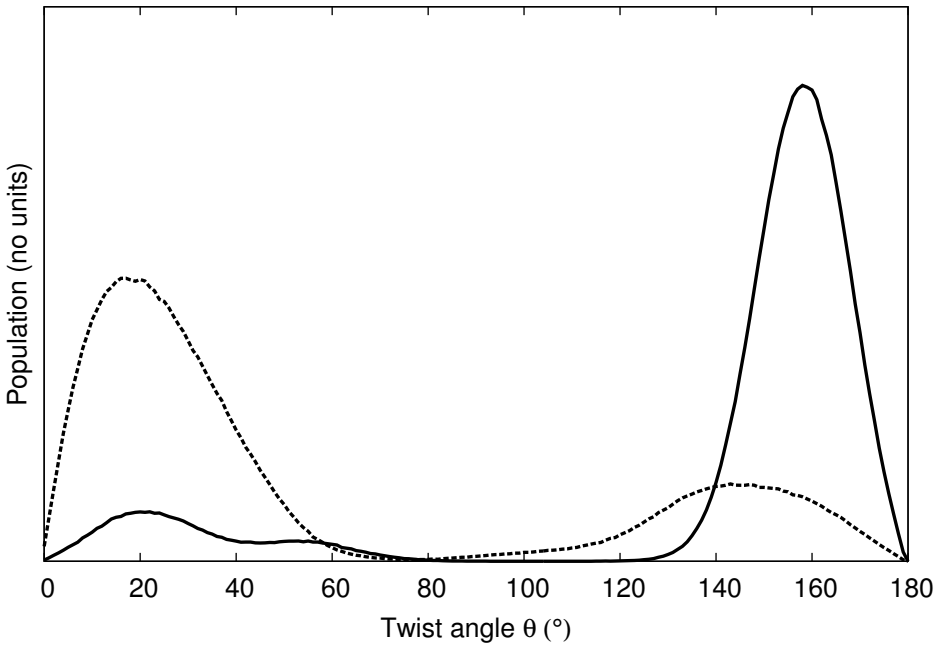

Figure 3.26: Histogram of twisting angle populations for Dye A layer structure (100 dye molecules at 30 wt.%) for neighbouring dyes (solid), and next neighbours (dashed).

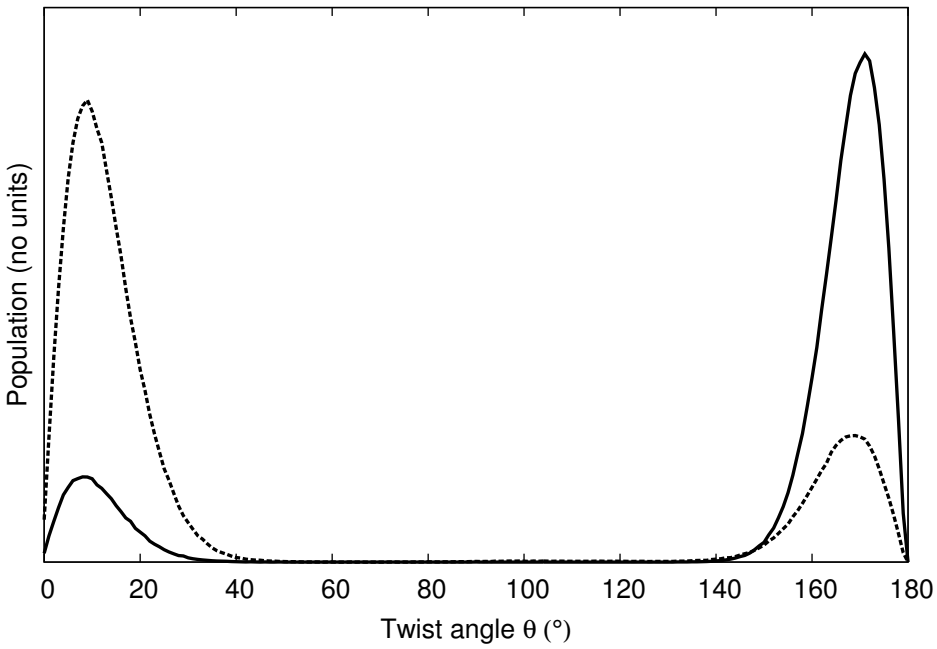

Figure 3.27: Histogram of twisting angle populations for Dye B layer structure (100 dye molecules at 30 wt.%) for neighbouring dyes (solid), and next neighbours (dashed).

with no chirality or persistent twist. The formation of a layer structure, where each aggregate is now in contact with another aggregate on either side, causes the columns to lose their chirality and adopt a non-chiral arrangement. Here, retention of a chiral column in a layer structure would lead to unfavourable interactions from sulphonate groups pointing into a neighbouring column. This provides a strong driving force for the columns to adopt an achiral formation.

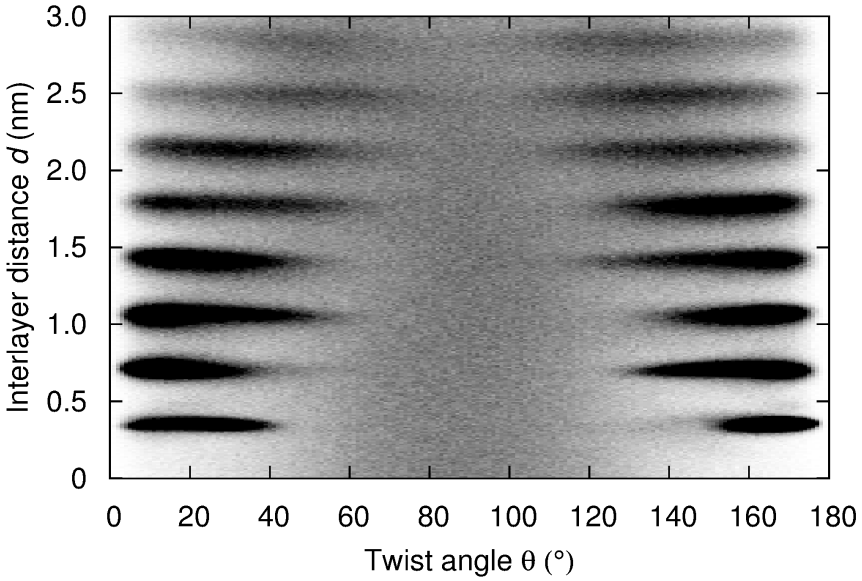

Figure 3.28: Histogram of twisting angle between all pairs and distance along the stack as a function of probability for the Dye A layer structure (100 dye molecules at 30 wt.%).

Column-column distance analysis

Analysis of the distance between columns within the layer for Dye A (fig 3.30) shows an average spacing of 1.75 nm. Previous experimental analysis,[3,4] using SAXS (Small angle x-ray scattering), for Dye A in the smectic region shows a 1.7-1.74 nm intra-layer spacing (ILS) reflection in the same plane as the 0.35 nm intermolecular spacing, indicating a repeat structure within the layer plane. The interpretation for this value is a repeat in structure every 5 rows in the layer (fig 3.2, (c)). Simulation

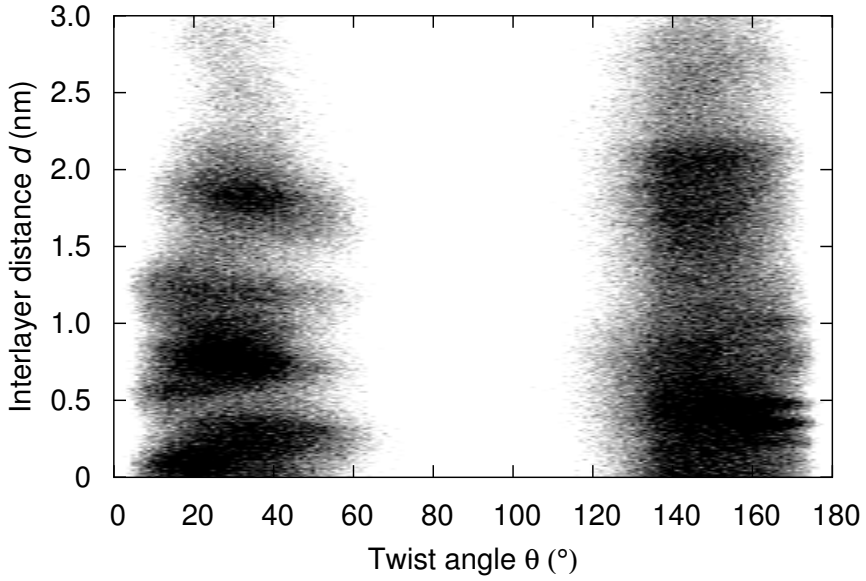

Figure 3.29: Histogram of twisting angle between all pairs and distance along the stack as a function of probability for the Dye B layer structure (100 dye molecules at 30 wt.%).

provides an alternative explanation for this peak, with the ILS reflection showing the distance between columns within the layer.

3.3.5 Discussion and conclusions

Extensive atomistic level molecular dynamics simulations, for Dye A and Dye B in aqueous solution have been used to provide a molecular level picture of the preferred stacking structure within chromonic aggregates. Results have shown that both systems have a predominant preference for a head-to-tail stacking arrangement, thus minimizing electrostatic repulsion between charged sulphonate moieties on neighboring dye molecules. Chirality is observed in aggregates for Dye A at low concentrations, at 9 and 20 wt.% solutions, despite the molecules themselves being achiral. All concentration systems for Dye B (9, 20 and 30 wt.% solutions) reveal aggregates with no chirality. The cause of chirality within Dye A aggregates can be deduced from the structural difference between Dye A and Dye B. Where Dye

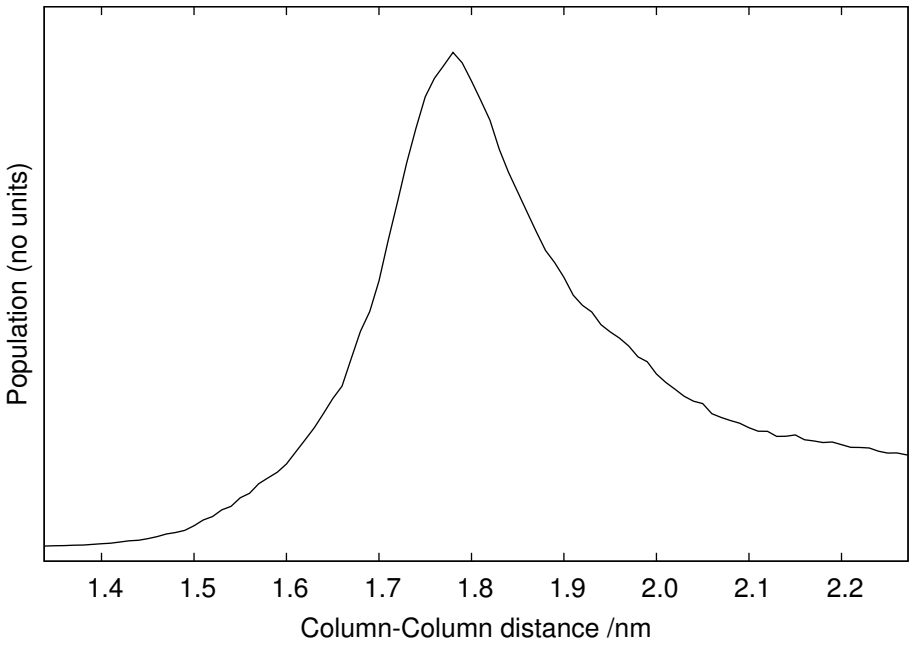

Figure 3.30: Preferred distance between neighbouring columns in the Dye A layer structure.

A has a methoxy group present, Dye B has a chlorine group, thus indicating that the methoxy group is crucial in causing chirality in these aggregates, or that the presence of chlorine in Dye B has a strong enough effect to disrupt potential chiral order. The driving force behind this is likely to be the size of the methoxy group. This causes adjacent dyes to lie at a persistent twist angle with respect to each other.

The cross sectional area for Dye A aggregates is shown to be concentration-independent, and independent of whether the aggregate has adopted a twisted structure or not. The computational CSAs successfully matches the experimental values, and thus allows a reinterpretation of the Dye A aggregate structure being a single width column, not a double width column.

A layer structure is observed in aggregates for Dye A at a higher concentration of 30 wt.% solution. The layer structure is formed by the alignment of columns in close proximity to form a layer. The aggregates are found to be non-chiral. The driving force for this can easily be deduced as, in an untwisted structure, the sulphonate groups do not point unfavorably into a neighbouring column.

Many structural features of the layer structure can be directly compared with

experimental evidence from X-ray measurements. Both models show a one molecule thick layer. Both models are also shown to be biaxial, however, with different ordering for the two systems. SAXS and WAXS reflections from the experimental literature data show a 0.35 nm repeat distance (corresponding to distance between neighbouring dye molecules in a stack), and an approximate 1.7-1.74 nm repeat distance (corresponding to a repeat every 5 layers). Both of these repeat distances have been observed in the computational layer structure, with 1.75 nm being observed as the distance between neighboring columns in the plane of the layer. This matches to experiment, providing for a reinterpretation of the model for the smectic layer structure in these systems.

The higher concentration simulations of Dye B, at 30 wt.%, show the formation of a uniaxial layer, where aggregates come together via side-to-side and face-to-side interactions. In principle, this layer structure has many similarities to the layered structure of Dye A; however Dye B shows the formation of smaller columns, with the ordering of these columns within the layer being isotropic. These two newly proposed layers mimic many aspects of the higher concentration structures suggested by Bottrill and Tiddy,[3,4] with the exception of molecules reoriented to favour sulphonate-solvent interactions.

The preservation of the column structures within the layers shows a more feasible packing model than the brickwork model suggested previously.[3,4] Both layer structures shown present a highly promising chromonic layer model, which is more energetically favourable than the literature model. The literature models suggest a highly energetically unfavorable configuration. The simulation findings show a layer structure formed without the breakdown of the columns, with the favorable aggregation properties observed at lower concentrations still being retained. The new model also shows a far simpler structure when considering a transition from the nematic phase to a smectic, with the alignment of columns along a plane excluding the need for an intercalated brickwork structure.

A range of absorption and linear dichroism spectroscopies, optical microscopy and cryogenic electron microscopy (cryo- TEM) have been used on cyanine dye systems in the literature.[23,24,94–98] Exotic structures, such as tubular architectures,[94]

threadlike arrangements,[95,96,98] fibrils[94] and complex fiber bundles[98] have all been suggested as packing models for various cyanine dyes. Another exotic model suggested is the hollow chimney column,[23,24] a complex brickwork structure where the feasibility of the packing arrangement has been heavily debated. The Dye A layer structure could be used to explain the formation of a hollow chimney column, with the layer curling up to form a tube (the layer is wrapped onto a cylindrical surface), a process already found experimentally for C8S3 dye molecules in water.[97]

Chapter 4

Atomistic studies of the Bordeaux dye

4.1 Introduction

Bordeaux dye, the product of the sulphonation of the *cis*-dibenzimidolazole derivative of naphthalenetetracarboxylic acid, is an anionic chromonic dye (Fig 4.1).[5]

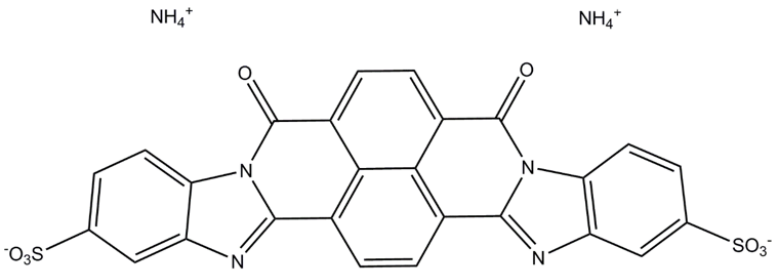

Figure 4.1: Molecular structure of the Bordeaux dye with ammonium counterions.

A wide variety of experimental techniques have been employed to study the behaviour of Bordeaux dye molecules in solution. Polarising Microscopy[5] revealed that nematic textures are observed at room temperature at a concentration of 6 wt.%, a value much lower than that found in many aggregating systems. X-ray scattering results[5] show that the compound self-assembles to form columnar aggregates, with a concentration-independent repeat distance of 0.34 nm between molecular planes,

typical of many chromonic systems. Typically, chromonic mesogens tend to favour a face-to-face aggregation (as discussed in Chapter 1) resulting in a cross sectional area of one molecule only. Bordeaux dye has been found to possess a cross sectional area of 2.5 times larger than that of a single molecule, suggesting an aggregate with two molecules within each molecular plane of the column.[5]

A schematic of the double-width column structure, as well as the two possible arrangements for the Bordeaux dye systems can be seen in fig 4.2. A third possible arrangement is where the sulphonate groups are in close proximity to each other has been eliminated as a possibility. Strong repulsion would occur between these sulphonate groups making this structure highly unfeasible.

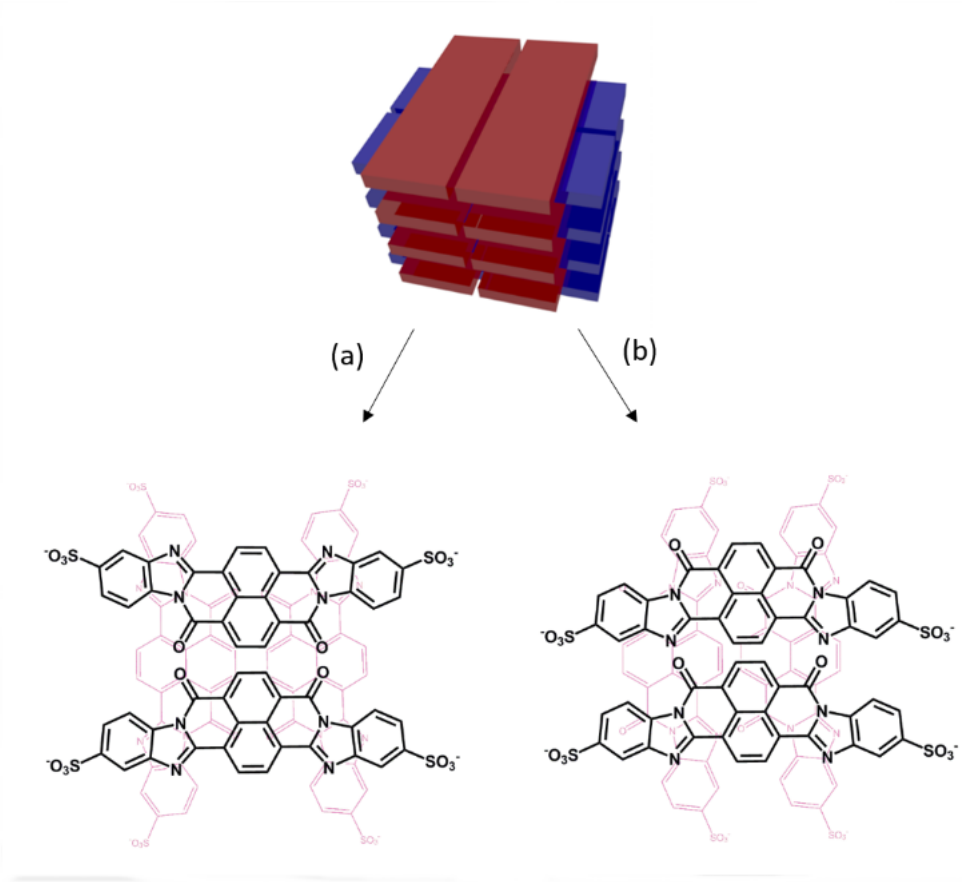

Figure 4.2: Proposed double-width column structure (top) with the two possible arrangements within the aggregate for minimal repulsion between sulphonate groups.

The low concentration onset of the nematic phase (6 wt.%) is unusual. In comparison, the well studied chromonic Edicol (sunset yellow) is typical of many stan-

dard chromonic systems, posessing a nematic onset of 30 wt.%,[43–46] considerably higher than the Bordeaux dye. A phase diagram mapping the phase behaviour of the system as a function of varying temperature and concentration can be seen in fig 4.3. The coexistence region is a two phase region where the system is transitioning from an isotropic to a nematic.

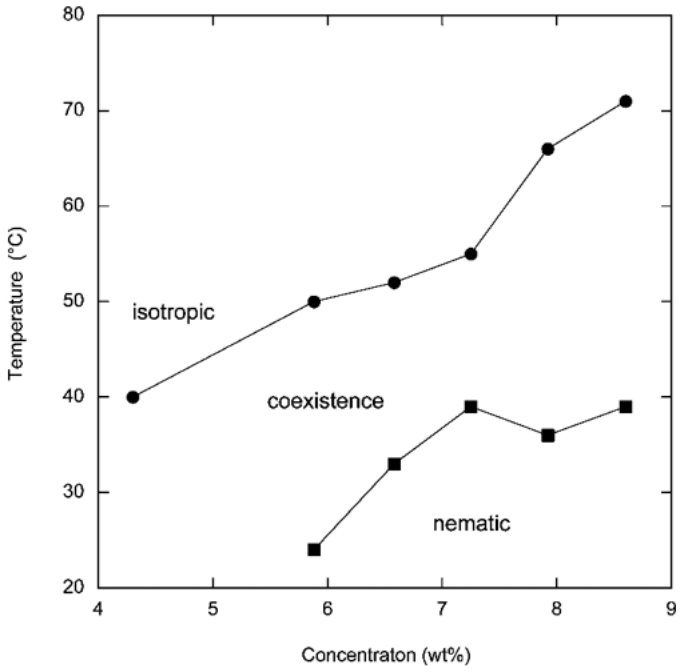

Figure 4.3: Phase diagram for Bordeaux dye. (Taken from Tomasik.[5])

The extensive polyaromatic ring structure, with no flexible linking groups, of Bordeaux dye results in a strong binding affinity between dye molecules, and thus the formation of liquid crystal phases at low concentrations. The large "stacking free energy change" (energy of the molecule in an aggregate relative to a molecule freely in solution) of 9.2 $k_{\mathrm{B}}\mathrm{T}$[5] (in comparison to the anionic azo dye Sunset Yellow value of 7.25 $k_{\mathrm{B}}\mathrm{T}$)[47] quantitatively demonstrates the strength of the binding affinity relative to a typical chromonic system. The theory[5] behind calculating the stacking free energy change from experimental data will be now discussed.

In all chromonic liquid crystal systems, the absorption spectrum changes as aggregation between molecules occurs. The two primary changes are a decrease in the absorption coefficient as aggregation increases, and a shift in wavelength of the absorption maximum as dye concentration increases.

One theory used to explain both of these changes in the absorption spectrum is the exciton coupling theory, i.e., coupling of transition moments of the constituent dye molecules. The non-covalent binding between dye molecules is taken into account by introducing off-diagonal elements into the Hamiltonian for the aggregate.[99] Making the coupling between neighbouring molecules constant, β, allows the energies ($E1$) and dipole moments ($P1$) of the eigenstates of the excited aggregate to be calculated, as shown in Equation 4.1 and 4.2, where j is the index for the excited states of an aggregate of i molecules

$$E_j = E_1 + 2\beta \cos\left(\frac{j\pi}{i+1}\right), \tag{4.1}$$

$$P_j = \sqrt{\frac{2}{i+1}} \sum_{k=1}^{i} P_1 \sin\left(\frac{jk\pi}{i+1}\right). \tag{4.2}$$

As the number of molecules in an aggregate increase, so does the number of eigenstates of the excited aggregate. The energies of these growing excited-state aggregates increasingly deviates from the excited-state energy of a single molecule, leading to a broadening of the absorption coefficient spectrum and a lowering of the peak absorption coefficient. The dipole moments of these additional states are also not equal, with the absorption spectrum shifting to either a higher or lower energy, as determined by the sign of the coupling term in the Hamiltonian.

A Lorentzian function, with a suitable width and coupling term, can be attributed to each eigenstate of the excited aggregate. As the aggregate grows, the result is that the maximum absorption coefficient shifts in wavelength and decreases in magnitude.

Whilst it is impossible to fit the data exactly to theory, due to the complexity of absorption spectra being higher than coupled exciton theory predicts, the decrease in the absorption coefficient at the wavelength of maximum absorption and at the wavelength of the single-molecule absorption maximum are qualitatively predicted. Both of these changes can be modeled by equation 4.3, whereby the assumption is that the absorption coefficient for extremely large aggregates a_∞ can be used as a fitting parameter in the relationship between the absorption coefficient of an

aggregate with i molecules

$$a_i = a_1 + (a_\infty - a_1)\cos\left(\frac{\pi}{i+1}\right), \tag{4.3}$$

where a_1 is the absorption coefficient for a single molecule. The aggregate sizes and distributions calculated from the UV-vis spectra can be then used in the following aggregation theory equations utilized to analyse X-ray scattering data

$$X_i = i\left(\frac{L^3}{v}\exp\alpha\right)^{i-1} X_1^i \tag{4.4}$$

where X_i is the volume fraction of an aggregate of size i, v is the volume of a single molecule, L is the cube length, $\propto$ is the stacking free energy change in units of k_BT, and X_1 is the volume fraction for a single molecule, which is defined as

$$X_1 = \frac{(1+2\phi z) - \sqrt{1+4\phi z}}{2\phi z^2} \tag{4.5}$$

$$z = \frac{L^3 \exp^{\alpha}}{v} \tag{4.6}$$

thus yielding a stacking free energy $\propto$. A full derivation for the aggregation theory used to obtain the stacking free energy and volume fraction terms for 4.4 and 4.5 can be found in the Tomasik and Collings paper.[5]

The aim of this chapter is to use atomistic simulation to probe the aggregation process of the Bordeaux dye in aqueous solution. Simulation will provide a molecular picture of the structure within an aggregate, shining a light on the exotic double-width column structure proposed. A thorough thermodynamic study will also be performed to examine the binding energy between dye molecules.

4.2 Computational details

The potential functions and interaction parameters were employed by the GAFF[60,80] parameters (described in section 3.2) as implemented using the simulation suite Gromacs.[81–85] The partial charges were calculated from BCC partitioning with the

AMBER **sqm** programme. The simulations were initally equilibrated in a canonical (constant-NVT) ensemble and then followed by extensive equilibration runs using the isobaric-isothermal (constant-NPT) ensemble. All simulations, unless otherwise stated, were performed at a temperature of 300 K, with a Nosé-Hoover thermostat[65] being employed. Pressure was controlled with a Parrinello-Rahman barostat[69] at atmospheric pressure and an isotropic pressure coupling. All simulations were initialized without bond constraints, until the simulation settled, bond constraints were then applied using the LINCS algorithm[89] and the time step was increased from 1 fs to 2 fs. The interaction cut-off for Lennard-Jones interactions was 1.1 nm, while short range Coulombic interaction were truncated at 1.2 nm. The long range part of the electrostatic interactions was represented by employing a Particle Mesh Ewald (PME) summation.[55] The TIP3P[90] water model was chosen, as this model is the most applicable to the GAFF forcefield.[91]

Initial simulations consisted of a system of 8 dye molecules and 16 counterion molecules randomly oriented and positioned in a cubic box. This process was followed by solvation of the system to give 4 wt.% concentration. The same process was applied for concentrations of 8 wt.% for a system of 30 dye molecules and 60 counterion molecules. For simplicity, any weight fraction concentration refers to the weight of both the dye and the counterions combined. Larger systems with 30 dye molecules showed the formation of many small stacks, with diffusion of these aggregates in solution being slow. Evidence from the 4 wt.% 8 dye molecule system indicated the formation of a single stack in solution if allowed to run for a longer time period. It is for these reasons that for analytical purposes, two single seeded columns of 30 dye molecules at 8 wt.% were created; one with full antiparallel arrangement between dye molecules and one with full parallel arrangement. Dye molecules within the pre-assembled stack were spaced at 0.35 nm intervals. These systems were then allowed to anneal over many hundreds of nanoseconds, to probe the stability of both configurations.

Two seeded double width column systems were pre-assembled to give idealised versions of the two suggested models[5] in the literature (discussed previously in section 4.1). Both systems are at a 23.4 wt.%, with the double width stack composed of

32 dye molecules. 64 counterion molecules are randomly oriented and positioned in the box, prior to the system being solvated. A further two systems were made with the same two seeded double width column structures, but with an ethyl ammonium counterion replacing the ammonium ions. This was then solvated to produce a 25.6 wt.% dye concentration.

The free energy of association (using a PMF, as discussed in Chapter 2) was calculated for a dimer, trimer and tetramer system at 3 temperatures: 300, 310 and 320 K. All systems were run at a concentration of 2 wt.%. A larger simulation cell was required to provide adequate space to pull the molecule far enough to where the molecules are no longer interacting. The simulations were carried out using the pull code implemented in the Gromacs package.[81] Initial configurations were generated with constrained distances between the centre of masses of a molecule at the end of a stack, and the centre of mass of the second molecule in the stack. For systems with more than 2 molecules (trimer and tetramer), the centre of mass for the molecules not being pulled are calculated as a single unit. This explains why the binding energy profiles for larger systems are increasingly shifted with respect to distance. A pull rate of 0.01 nm ps^{-1} was employed. For each chosen distance between centre of masses, the system was equilibrated for 1 ns, followed by a 50 ns production run where forces were outputted every 10 fs, for a simulation of 2.5 x 10^7 steps with a timestep of 2 fs. The distance between the centre of mass was varied from 0.3 to 2 nm, where spacing between the neighbouring points varied between 0.01 to 0.1, with smaller spacings for regions near the global minimum, and larger spacings as the dye molecule moves further out into solution.

Various analysis methods were used to quantitatively assess the structure of the aggregates. Details of the theory and equations used for each can be seen in chapter 2. To use these analysis techniques (such as the interlayer distance, twisting angle, distance-angle pair correlation functions, column-column distance), various atoms on the molecule are selected. To calculate the unit vector defining the normal to the plane of the molecule, atoms C6, C7 and C8 were chosen (as highlighted in green on fig 4.4). These atoms are heavily imbedded in the central aromatic system, and therefore retain a planar structure throughout. The unit direction vector, which is

defined by atoms C21 and C20, is highlighted in orange on fig 4.4, with an arrow marked to show the direction of this vector.

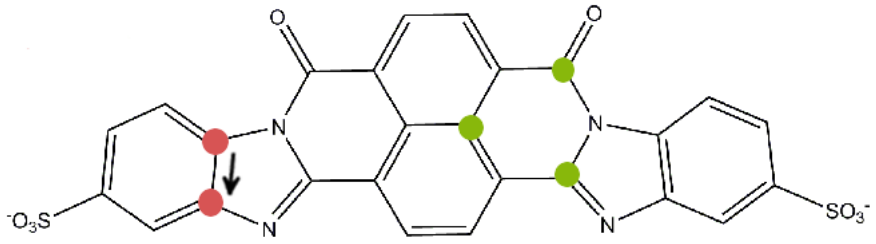

Figure 4.4: The Bordeaux dye molecule showing the atoms selected for analysis techniques, where the orange atoms (C21, C20) show the atoms used for the unit direction vector, and the green atoms (C6, C7, C8) are used for the unit vectors defining the normal to the plane of the molecule.

4.3 Results

4.3.1 Aggregation behaviour and short range order

An initial small system of 8 dye molecules in aqueous solution (4 wt.%) was simulated to inspect the aggregation process of the Bordeaux dye in solution. Starting from a randomly dispersed and oriented system, molecular aggregation of the dye molecules occurs quickly, with a dimer and trimer forming in under a nanosecond. Initial inspection shows no preference for a parallel (where sulphonate groups are on the same side) or antiparallel arrangement (where the sulphonate groups are on opposing sides respectively) between neighbouring dye molecules (Figure 4.5). The system evolves to form dimers and trimers, with these entities merging to form a hexamer and a separate dimer within 12.14 ns. Within approximately 200 ns, the two units meet to form a single stack. This stack shows a predominant preference for an anti-parallel arrangement between neighbouring dye molecules, and retains its structure for a further several hundred nanoseconds, and does not fragment, indicating an equilibrated structure.

The final equilibrated stack (Fig 4.6) shows face-to-face aggregation of dye molecules. Short range order between neighbouring dye molecules shows a preference for an

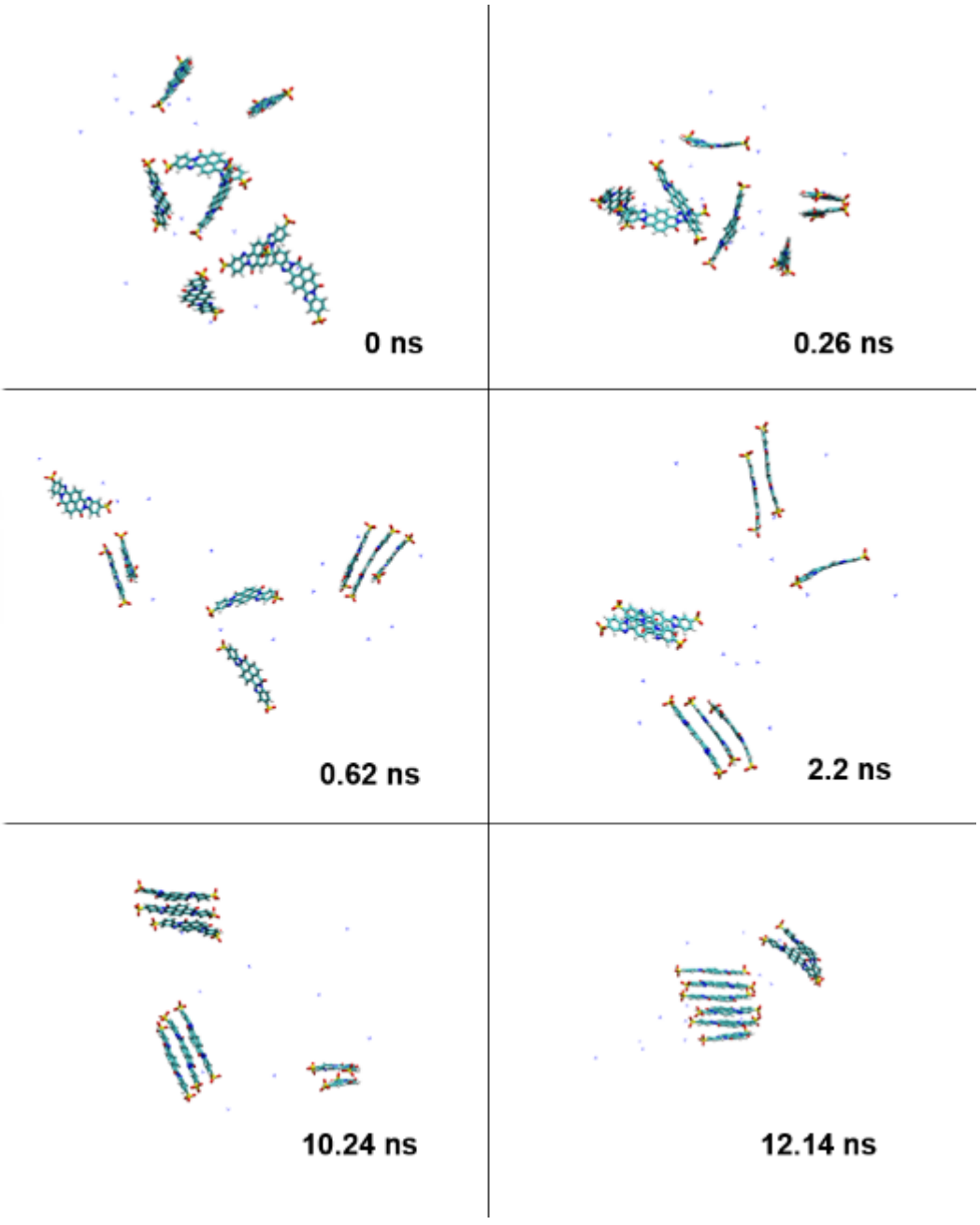

Figure 4.5: Aggregation process of the Bordeaux dye in solution (water has been removed from the snapshots for visual clarity). The sulphonate groups are the large red and yellow unit on either long end of the molecule.

antiparallel configuration, where sulphonate groups alternate in direction along the stack, thus minimising electrostatic repulsion that would occur with a closer proximity arrangement. A single pair of neighbouring dye molecules possess a parallel configuration within the stack, and are retained within the structure for the duration of the simulation (several hundreds of nanoseconds).

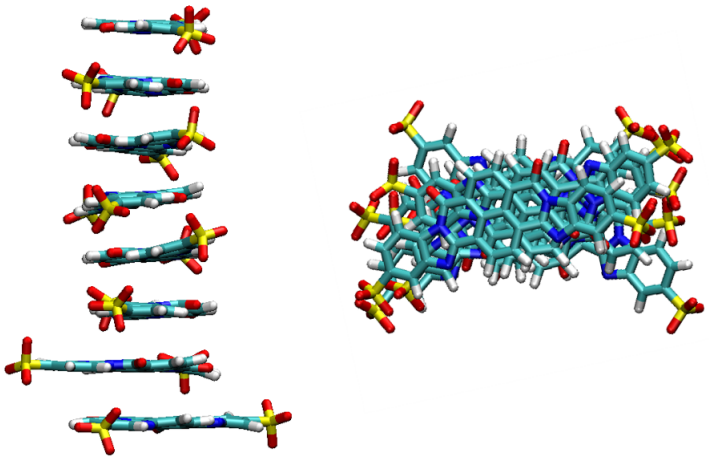

Figure 4.6: Short range order within a Bordeaux dye column, side view (left) and sky view down the column (right).

Free energy of association

The free energy of association (or "stacking free energy change") for a molecule in an aggregate relative to a molecule in solution can be obtained for a dimer, trimer and tetramer system from PMF curves at 300 K in fig 4.7. The aggregation process is shown to be isodesmic, with approximately the same energy increment occurring for the addition of each new molecule to a column, independent of the size of the column (table 4.1). The free energy profile for each system shows the presence of two potential wells: a global minimum (ΔG_{min}) and, most noticeable in the dimer system, the presence of a second local minimum (ΔG_{trans}). Visual inspection of the dimer system shows two different modes of packing at each respective potential well (fig 4.8). The global minimum (a) shows a H-aggregate, whereby the stacking is untilted, and the molecules have adopted an antiparallel configuration with respect to the sulphonate regions. The local minimum (b) shows a J-aggregate, whereby the

System	ΔG /kJmol^{-1}	ΔG /k_BT	ΔG_{trans} /kJmol^{-1}	ΔG_{trans} /k_BT
Dimer	67.3	27.1	26.2	10.6
Trimer	67.9	27.4	28.4	11.4
Tetramer	67.8	27.3	25.6	10.3

Table 4.1: Free energy of association for dimer, trimer and tetramer at 300 K

stacking is off-set, and the molecules have also adopted an antiparallel configuration with respect to the sulphonate regions.

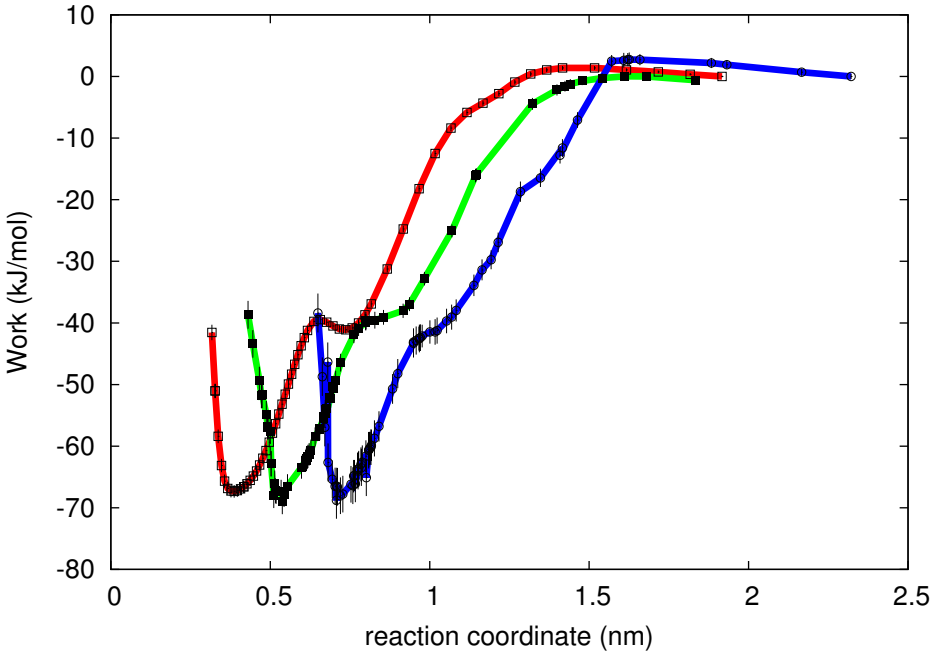

Figure 4.7: Free energy of association for a dimer (red), trimer (green), and tetramer (blue) at 300 K. Error bars for each data point are shown in black.

For discussion purposes, and comparison to experimental results, the values for the dimer system will be used as the computational reference. Here, as the system is isodesmic, all three wells have the same depth. Compared to the experimental value of 9.2 k_BT,[5] the stacking free energy change extracted from simulations is markedly higher, exhibiting a global binding energy (ΔG_{min}) of 27.1 k_BT (approximately 3 times larger). The presence of a second well (ΔG_{trans}), which is the configuration change from a H-aggregate to a J-aggregate (Fig 4.8), has a ΔG value of 10.6 k_BT. This value matches strongly to the experimental value found, giving credence to the theory that the experimental value probes from a global minimum to another local minimum, and not to a fully dissociated dye in solution. The difference in values

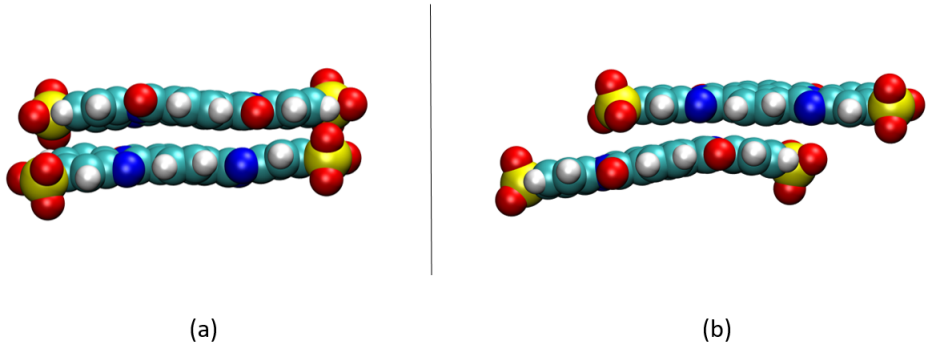

Figure 4.8: Visual inspection of the two potential wells in the dimer system, where (a) is the H aggregate at the global minimum, and (b) is the J aggregate at the local minimum.

can be explained by several assumptions made in the experimental derivation of the free energy of association.

Simulations have shown that Bordeaux dye aggregates are far more complex than simple rods, and can adopt more configurations than the predominantly face-to-face aggregation seen in most chromonic systems. The presence of different packing modes, such as the shifted junction (J-aggregate), and hence the possibility of further intra-aggregate arrangements such as the Y-junction, produce a far more complex system than with the oversimplified rod shape normally used for chromonics. It should be noted that the complex Y-junction packing mode has been experimentally observed for several block copolymer and surfactant systems in the literature.[6] A schematic of the N phase in a chromonic liquid crystal for both the standard model, and the more complex model can be seen in Figure 4.9.

The consequences of this more complex structure is that the correlation lengths ξ_L (measured along the stacking distance) of the aggregate are different between the standard and complex model. A standard model has a correlation length corresponding to the entire aggregate, whereas the complex model has two spatial scales: a short scale, where ξ_L is the locally correlated assembly of dye molecules, and ξ_{L-tot} is the summation of all of these smaller units to form the whole aggregate (fig 4.9). Experimentally, X-ray is unable to observe the whole aggregate composed of these smaller units, therefore not reflecting the true dimension of the aggregates. The size and distributions of aggregates is crucial in the calculation of the stacking free energy change (see section 4.1), therefore the failure of X-ray to correctly distinguish

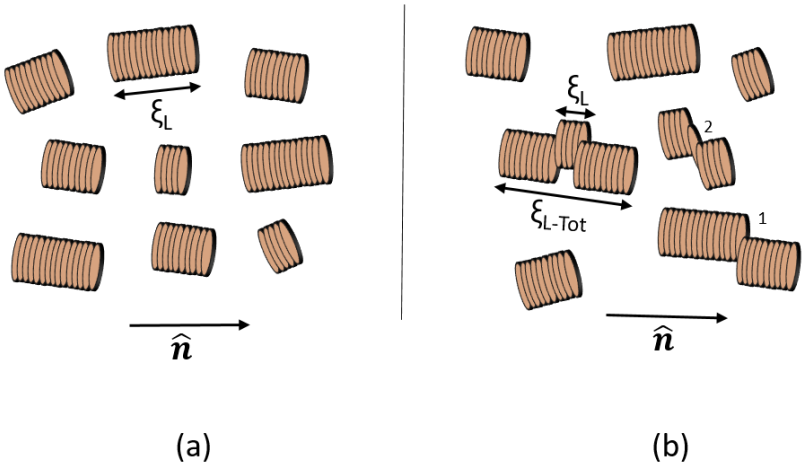

Figure 4.9: Schematic models of the N phase in a chromonic liquid crystal; (a) standard model with rod-like aggregates, (b) a model with shifted configurations, such as shift junctions (b1) and Y-junctions (b2). Picture reproduced from Park *et al.*[6]

between the two spatial scales in a complex system invalidates the assumptions that go into the calculation of a binding energy from experiment.[5]

The presence of a more complex nematic structure, due to shifted configurations and the periphery charged groups, provides additional evidence that ionic chromonic systems deviate from the simple hard rod theories presented by Onsager,[26] which is commonly used to describe these systems.

The applicability of the Onsager theory can be examined using the volume fraction together with the length-to-width ratio (of the molecule), whereby the product of these for a nematic-isotropic transition should be approximately 4).[100,101] In the literature, several chromonic dyes have a value well below 4, with 1 for disodium chromoglycate, 1.3 for Sunset Yellow and 0.2 for Bordeaux dye.[5] Experimentally, X-ray leads to a gross underestimation of the aggregate lengths, due to the inability to discern the connected branches formed by shifted configurations and Y-junctions within an aggregate. The complex structure gives further credence to its existence by explaining why the nematic order is so high for that volume fraction, and why there is such a disparity between the nematic onset values found, and the conditions required by the Onsager theory.

Free energy of association with varying temperature

The free energy of association can be broken down into two components: enthalpy and entropy. Thermodynamic analysis can be performed to separate the enthalpic and entropic contribution, thereby helping to gain insight into the driving forces for aggregation. To obtain the enthalpic and entropic components, the free energy of association has been calculated for a dimer, trimer and tetramer at 3 different temperatures: 300, 310, 320 K.

From the definition of the Gibbs free energy in a standard state

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ, \tag{4.7}$$

where ΔH° is the enthalpy and ΔS° is the entropy. From ΔG° it is possible to calculate the equilibrium constant K for the standard state

$$\Delta G^\circ = -RT \ln K, \tag{4.8}$$

where R is gas constant. A combination of (4.7) and (4.8) yields

$$\ln K = \frac{-\Delta H^\circ}{RT} + \frac{\Delta S^\circ}{R}. \tag{4.9}$$

Plotting $\ln K$ as a function of $1/T$ allows the slope and intercept to provide the enthalpic and entropic contributions to associations. The linear Van't Hoff plot for the global free energy change, ΔG°, can be seen in figure 4.10, and the plot for $\Delta G^\circ_{\text{trans}}$ can be seen in figure 4.11. Data extracted from the trend lines for each system to achieve enthalpic and entropic contributions are shown in table 4.2.

Thermodynamic property	Dimer	Trimer	Tetramer
$\Delta H^\circ_{\text{global}}$/kJmol^{-1}	-95.4 ± 20.2	-83.1 ± 36.1	-124.1 ± 18.4
$\Delta S^\circ_{\text{global}}$/Jmol^{-1}K^{-1}	-92.4 ± 65.3	-48.9 ± 116.7	-191.8 ± 59.5
$\Delta H^\circ_{\text{trans}}$/kJmol^{-1}	-36.6 ± 4.5	-33.0 ± 3.7	-50.6 ± 6.4
$\Delta S^\circ_{\text{trans}}$/Jmol^{-1}K^{-1}	-34.5 ± 14.4	-13.9 ± 11.8	-81.3 ± 20.5

Table 4.2: The enthalpic and entropic values for a dimer, trimer and tetramer system, as calculated from the trend line on the linear Van't Hoff plots of figures 4.10 and 4.11.

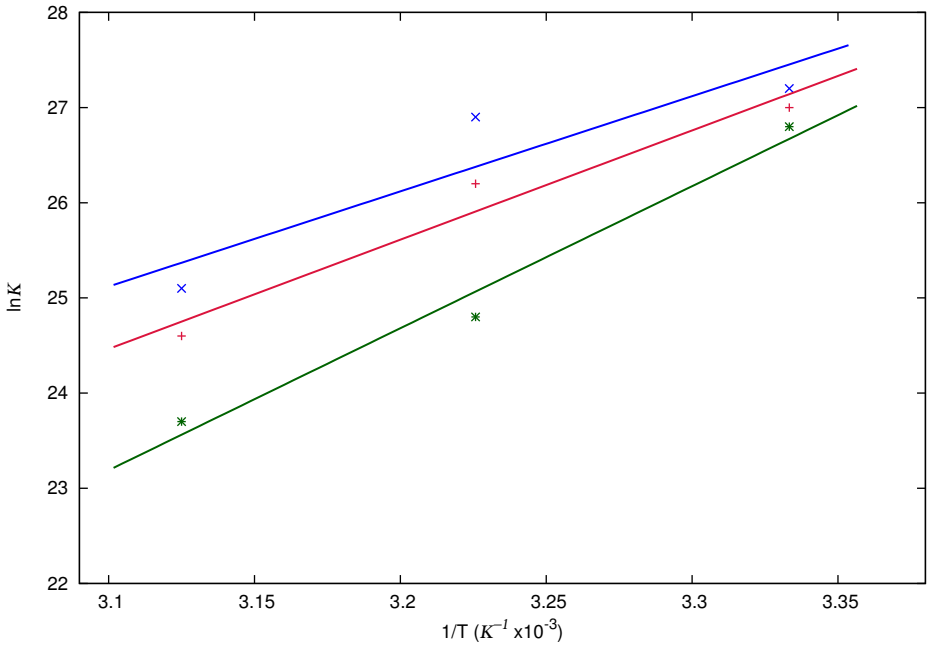

Figure 4.10: Van't Hoff plot for ΔG° showing $\ln K$ as a function of $1/T$for a dimer (red), trimer (blue) and tetramer (green).

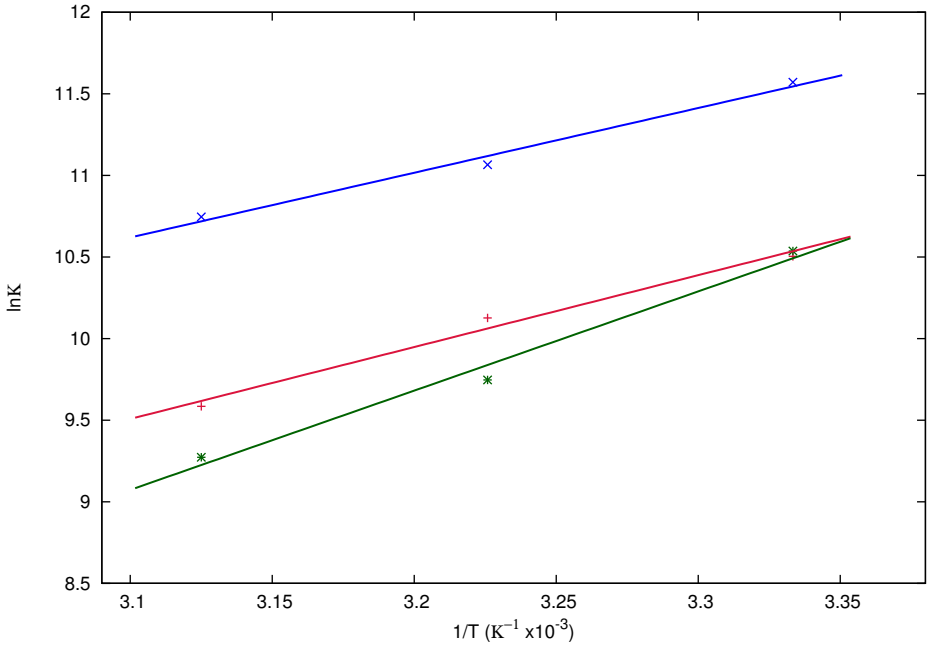

Figure 4.11: Van't Hoff plot for $\Delta G^{\circ}{}_{\text{trans}}$ showing $\ln K$ as a function of $1/T$ for a dimer (red), trimer (blue) and tetramer (green).

The thermodynamic analysis for the global free energy (ΔG°) for a dimer, trimer and tetramer reveal that the aggregation process is enthalpically driven. The $\Delta H^{\circ}_{\text{global}}$ values are more negative than the corresponding ΔG° values, showing that aggregation is completely governed by the attraction between dye molecules in solution. Aggregation is unfavourable with respect to entropy, as can be highlighted by the negative $\Delta S^{\circ}_{\text{global}}$ values, which lowers the overall free energy of the process. This is in contrast to many conventional lyotropic systems, where formation of micelles is often entropically driven by release of weakly bound water molecules.[20] The thermodynamic analysis for the change in configuration from the global minimum to a local minimum ($\Delta G^{\circ}_{\text{trans}}$) for a dimer, trimer and tetramer also revealed an enthalpically driven process. The change in configuration from a J-aggregate to a H-aggregate is entropically unfavourable (as shown by the negative $\Delta S^{\circ}_{\text{trans}}$), but is outweighed by the enthalpic drive of adopting this configuration (a large negative $\Delta H^{\circ}_{\text{trans}}$).

However, we note several key caveats that should be highlighted in the process of attempting to separate out the enthalpic and entropic components of the association free energy. The dimer system has a very clear second well for all temperatures, so measurement of $\Delta G^{\circ}_{\text{trans}}$ can be made with high accuracy. However, the trimer and tetramer system proved more troublesome. Additional molecules in the stack, which are not present in a dimer system, appear to be able to interact with the molecule being pulled. Whilst an obvious change in the free energy is noted for all systems at the same free energy increment from global minimum to local minimum (second well), these additional interactions lead to the local minimum becoming more of a semi-flat line than a deep well. To calculate the $\Delta G^{\circ}_{\text{trans}}$, the value in the middle of flat region within the local minimum was therefore selected as the value to use. However, this makes the value both ambiguous (due to human error in selection), and potentially unreliable due to the $\Delta G^{\circ}_{\text{trans}}$ for those systems being potentially overestimated, as the presence of a true well would lower the $\Delta G^{\circ}_{\text{trans}}$ by approximately 1.5 kJmol^{-1} in comparison to the dimer systems.

4.3.2 Long range order within aggregates

Initial inspection of two larger systems (30 dye molecule systems both at 8 wt.%), showed identical behaviour to the smaller system, with self-assembly between the dye molecules occurring quickly to form dimers/trimers/tetramers and so on. Unfortunately, due to the size of these systems, diffusion of the small aggregates in solution to meet and form a single column require an extensive amount of simulation time. To overcome this issue, two pre-assembled columns of 30 dye molecules (at 8 wt.%) were constructed, one with a fully antiparallel arrangement (system A) between neighbouring dye molecules in the column, and the other possessing only parallel arrangements (system B). Both systems were solvated and allowed to evolve over several hundreds of nanoseconds.

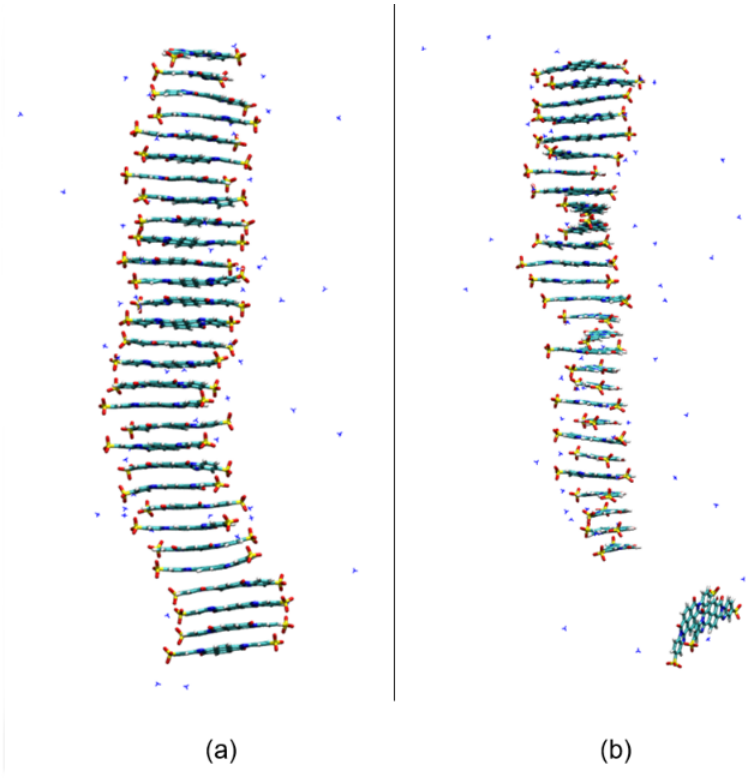

Figure 4.12: Long range order within a Bordeaux dye aggregate at 8 wt.%. Antiparallel System A (a), parallel System B (b). Counterions are highlighted in blue.

Visual inspection of both systems (Fig 4.12) shows that both aggregates retain a face-to-face arrangement of dye molecules. System A (Fig 4.12, a) retains a predominantly antiparallel arrangement between neighbouring dye molecules (with the

exception of one pair on the end of the column that has adopted a parallel configuration). System B (Fig 4.12, b) retains a predominantly parallel arrangement between neighbours, and appears to adopt a wider range of twisting angles between neighbours than System A, possibly to accommodate the non-ideal packing configuration. The fragmentation of a dimer unit from the main column unit is also observed for system B.

Interlayer distance

The interlayer distance is the average distance between neighbouring dye molecules in a stack. Analysis of both system A and system B show the same average interlayer distance of 0.34 nm (Fig 4.13). Experimental studies of various chromonic systems show a typical 0.34 nm peak corresponding to the repeat distance between molecular planes, thus matching the interlayer distance found for the simulated Bordeaux dye aggregates.

The interlayer distance cannot be used to distinguish between a parallel or antiparallel arrangement between dye molecules by comparison to experiment, as both configurations have the same preferred value; with only a small difference in the spread of values between system A and system B.

Analysis of twisting angle

Analysis of twisting angle between adjacent and next-adjacent molecules was used to observe, on average, how much molecules twist with respect to their nearest and next-nearest-neighbours for system A and B (Fig 4.14).

For neighbouring dye molecules, system A shows a sharp peak in the 160-180° region, with a preferred twist angle in the range 174-177°, corresponding to an antiparallel configuration. A small peak is also observed at 32°, owing to the presence of a single pair of parallel arrangement within the column. Hence, this shows that despite a predominantly antiparallel configuration, transition of a small fraction of neighbouring dye molecules to adopt a parallel arrangement is still energetically feasible. Analysis of next-neighbours shows a preferred twist angle value of 3-6°, con-

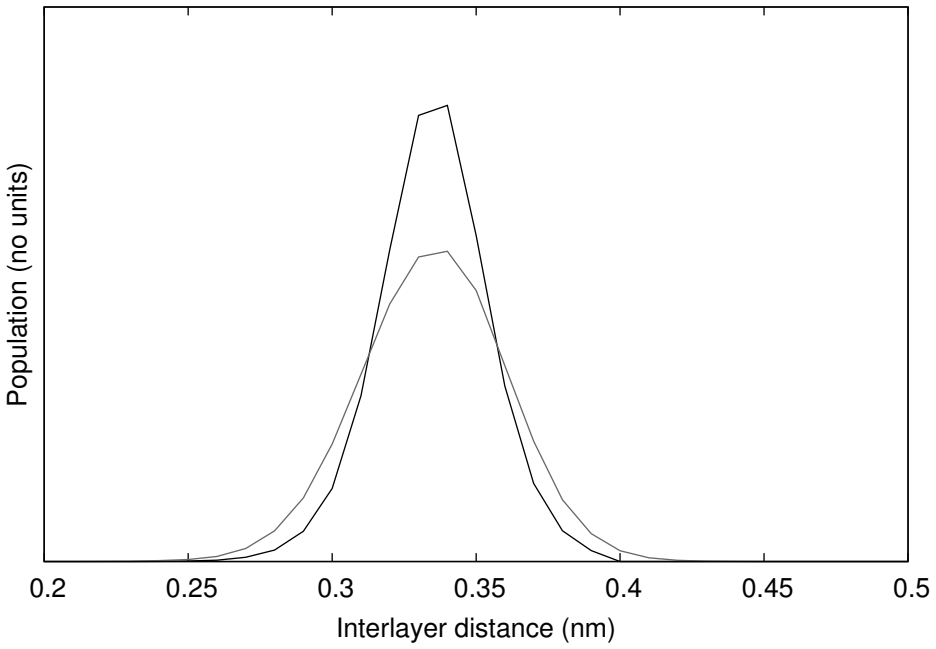

Figure 4.13: Interlayer distance between neighbouring molecules in a stack. (System A (black), system B (grey)).

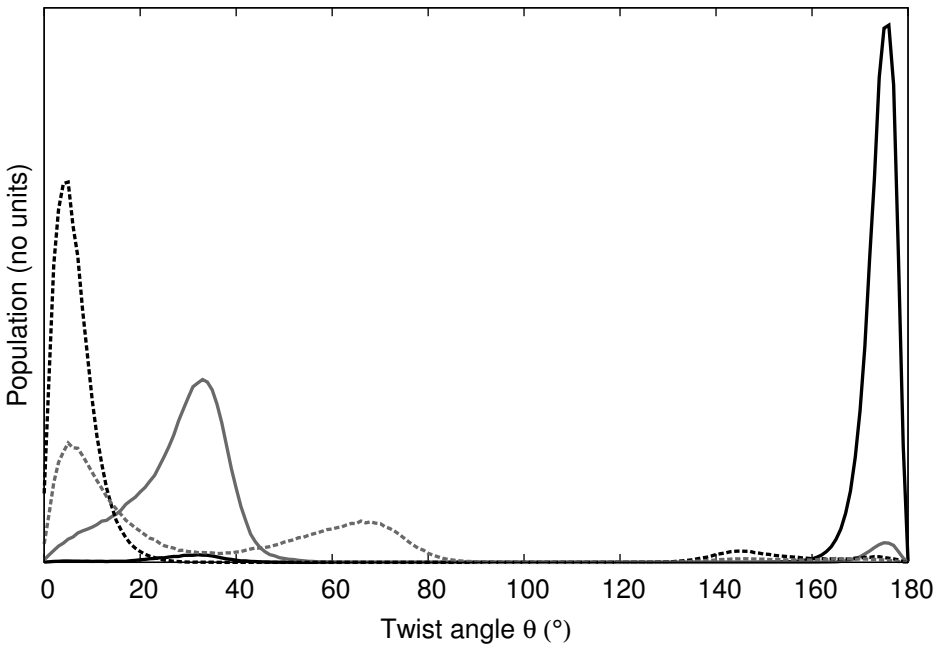

Figure 4.14: Twist angle for system A (black) and system B (grey), and next-neighbours system A (black dashed) and system B (grey dashed).

firming a parallel arrangement between next-nearest-neighbours within the system A aggregate.

For neighbouring dye molecules, system B shows a broad peak in the range 0-50°, with a preferred twist angle of 30-35°, corresponding to an off-set parallel configuration. Analysis of next-nearest-neighbours shows a broad range of twist angles from 0-80°, with a sharper peak at 4-8° and a smaller, broader peak at 65-70°. Both of these peaks correspond to an approximate off-set twist angle of 30° in either direction for next-nearest-neighbours.

Comparison of both systems show a marked difference in twisting angle, depending on whether the neighbouring molecules are in an antiparallel or parallel configuration. A parallel configuration adopts an off-set twist angle, with next-neighbours having the same off-set twist in either direction. An antiparallel configuration does not adopt an off-set twist angle.

Analysis of the twist angle between neighbouring dye molecules as a function of time (for an equilibrated system) allows for the dynamic nature of molecules within the aggregate to be studied (fig 4.15). Although all neighbouring pairs were calculated (approximately 15 pairs per system), a single antiparallel and parallel configuration were selected as typical representatives for both system A and system B, as shown in figure 4.15.

For neighbouring dye molecules in an antiparallel configuration, both System A and System B show a single configuration with an average twist angle of 170°. For neighbouring dye molecules in a parallel configuration, both System A and System B show the neighbouring pair can adopt one of two possible configurations. A twist angle of 10-15° and 30-35° are observed as the two possible configurations, with dynamic switching between these two configurations as the system evolves over time in an equilibrated state. A schematic for the two configurations can be seen in fig 4.16.

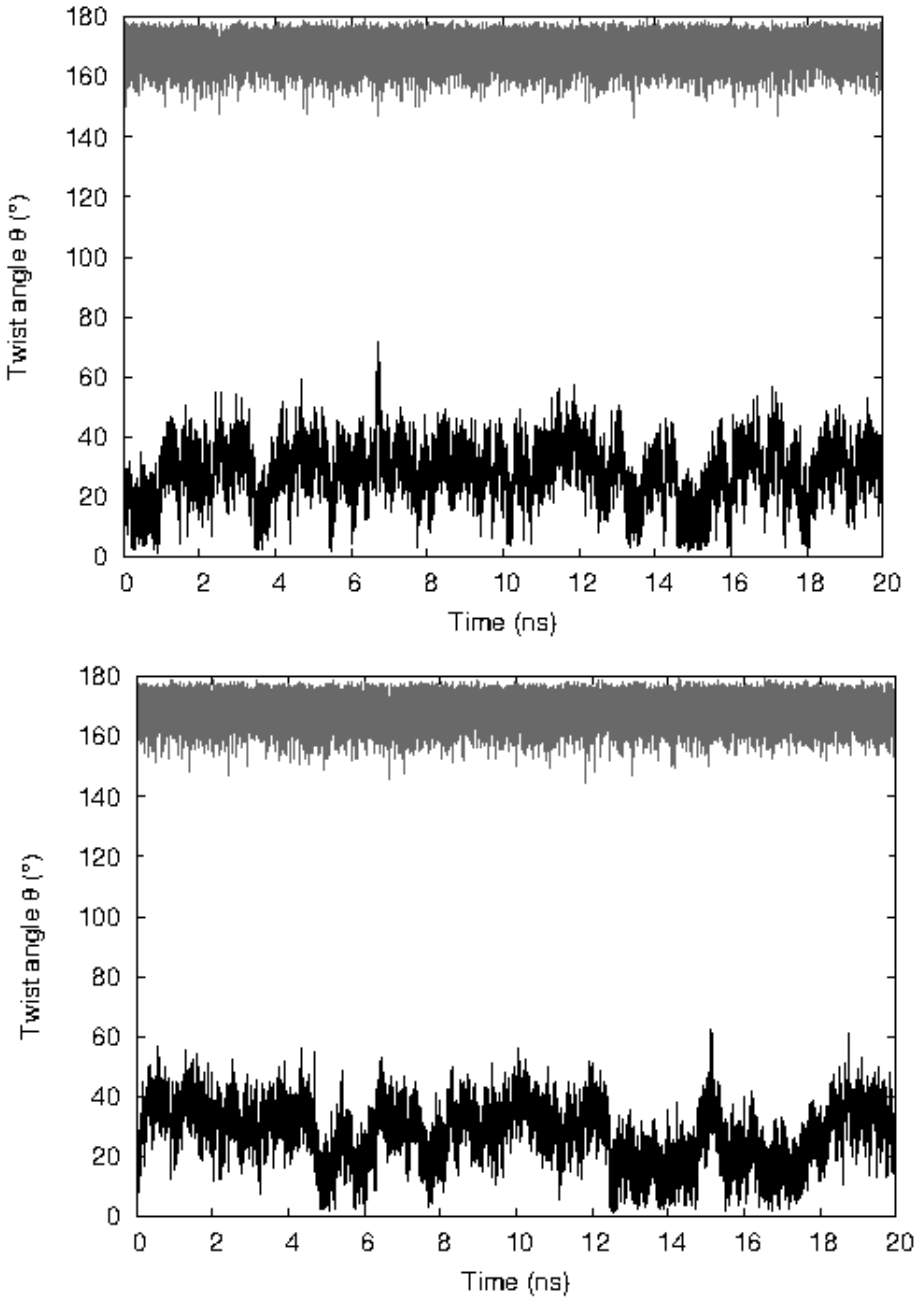

Figure 4.15: Time evolution of twist angle between neighbouring dye molecules in a stack for system A (top) and system B (bottom). Two typical cases for an antiparallel and parallel arrangement between neighbours have been selected for both systems.

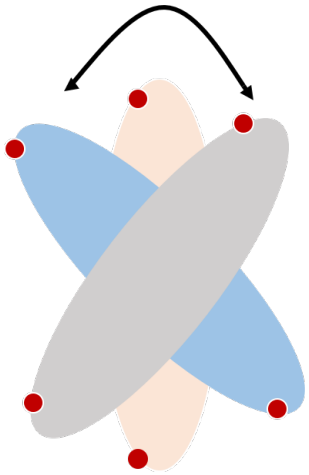

Figure 4.16: Schematic to show the two configurations (as shown by the grey and blue molecule) adopted when neighbouring dye molecules are in a parallel arrangement. Red dots indicate the charged sulphonate groups on a dye molecule.

Long-range twist angle analysis

To analyse the long-range order of both systems, pair correlation functions mapping the twist angle and distance as a function of occurrence were calculated (Fig 4.17). This allowed for any helical nature or long range twisting to be analyzed if present.

System A shows a single distribution of peaks at each packing distance (increments of 0.34 nm). Antiparallel stacking for neighbours is found at distances of multiples of 0.34 nm, while parallel arrangements from next neighbours are observed at distances of multiples of 0.68 nm. No helical or long range twisting behaviour is observed for system A.

System B shows either a single or multitude of peaks at each packing distance (increments of 0.34 nm). The bimodal distribution of peaks in many regions corresponds to the off-set angle the dye molecules adopt, with no preference shown for which direction the off-set occurs. No helical or long range twisting behaviour (such as a correlated off-set twist angle leading to a chiral column as seen in chapter 3) is observed for system B.

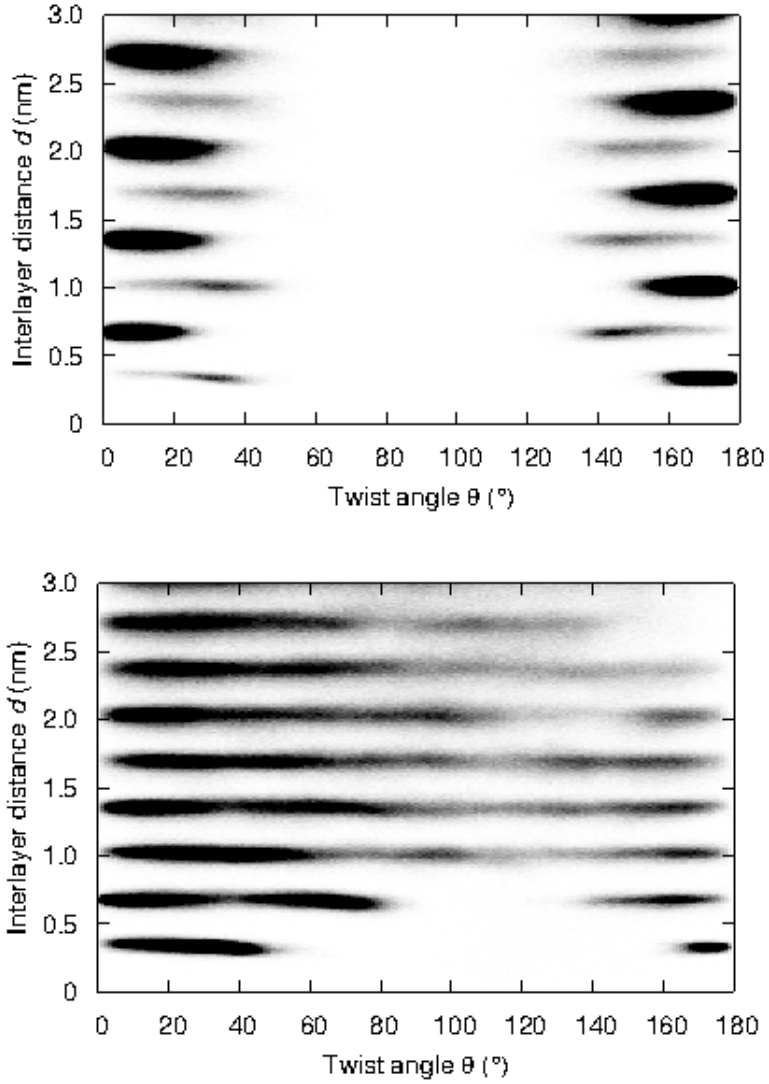

Figure 4.17: Histogram showing the population of twisting angle between all pairs and distances along the stack for system A (top) and system B (bottom).

Cross sectional area analysis

The cross sectional area (CSA) for both the system A and system B aggregates can be seen in table 4.3. Here, the method of calculation described in chapter 2 has been employed.

The value of a single molecule in the literature,[5] as calculated by the software package Gaussian, obtains an area of 128 Å^2. However, a simple calculation, whereby the length and width of a molecule are measured (fig 4.18), using the molecular graphics software Visual Molecular Dynamics (VMD) obtains an approximate area of 133 Å^2. The VMD calculation assumes a minimal sized rectangle around the molecule. The similarity between the two values confirms that the reported literature value corresponds to a single molecule where the surface accessible area has not been included. The software package Pymol allows an accessible surface area for a single Bordeaux dye molecule to be calculated, whereby a solvent probe of radius 0.14 nm is chosen to match the radius of a water molecule. Running this probe around a single molecule, and then dividing by two, gives a value of approximately 215 Å^2 for the solvent accessible surface area, which will be now used as the single molecule reference area, as this corresponds to the most realistic measure of surface area.

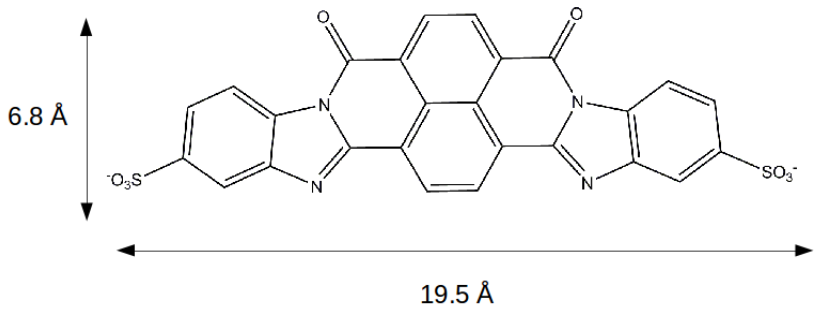

Figure 4.18: Dimensions of a single dye molecule obtained using molecular graphics software VMD.

System	Cross sectional area /Å^2	Error /Å^2
System A	211.93	1.66
System B	205.42	1.72
Experimental	324	4

Table 4.3: Cross sectional areas for aggregates in system A, system B and the experimental value.

In comparison to the area of one single molecule, system A and system B both possess a CSA corresponding to one molecule only (as shown in table 4.3). In contrast, the experimental value results in an aggregate CSA of approximately 1.5 molecules. Whilst this value excludes a double width column, it is significantly larger than a single width column, so must reflect a complex structure between these two column types.

Summary of system A and system B

An atomistic simulation study of the Bordeaux dye in aqueous solution has revealed that the chromonic mesogens aggregate in a face-to-face fashion to form single width stacks. A concentration-independent stacking distance of 0.34 nm is observed between neighbouring dye molecules in the aggregate. Spontaneous aggregation leads to a predominantly antiparallel arrangement between neighbouring dye molecules. However, pre-arranged stacks with a parallel arrangement largely retain this structure, indicating that both arrangements are energetically stable and there is likely a large barrier to rotation between the two stable arrangements. The twisting angle between neighbouring dye molecules in a parallel arrangement has been shown to be dynamic, with flips continually occuring between 10-15° and 30-35°, whilst the antiparallel arrangement has a single stable arrangement with a twist angle of 174-177°. The cross sectional area of these columns has been computationally calculated to be one molecule, with a more appropriate surface accessible area for a single molecule allowing a reintepretation of the experimental value to be 1.5 molecules. The new experimental value denounces the suggested double width column model, but indicates an aggregate structure more complex than that of a single width column.

A thermodynamic analysis of a range of different Bordeaux aggregate sizes at varying temperatures shows aggregation to be isodesmic, with similar energy increments occuring for the addition of each new molecule to the stack. The aggregation process is found to be driven purely by enthalpy, with the entropy contribution to association being unfavourable. The free energy profile shows the presence of two potential wells, a global minimum and a local minimum. The global minimum

corresponds to a stacking free energy of 27.1 k_{B}T, whereby the dye molecules are arranged as a H-aggregate. The local minimum corresponds to an offset arrangement (J-aggregate) relative to the lowest energy arrangement (H-aggregate) of 10.6 k_{B}T, whereby the dye molecules are arranged in an off-set configuration of a J-aggregate. The presence of two stable configurations shows a far more complex model of aggregation, whereby stacks could be made up of smaller units bridged by shift and Y-junctions. From an experimental view point, only these smaller units are identifiable, which suggests that an incorrect distribution of aggregates may have been identified in previous experimental work. Additionally, the experimental stacking free energy of 9.2 k_{B}T could be measuring the free energy change from the global minimum to the local minimum, not to where the molecule is freely dissociated in solution.

The recalculation of the surface accessible area for a single dye molecule allows the experimental cross sectional area for aggregates to be that of 1.5 molecules, not 2.5. This value excludes the double width column suggested by the experimental literature,[5] but is commensurate with the computational findings of a predominantly single width column with shifted and Y-junctions, which would lead to a larger solvent accessible surface area.

4.3.3 Double-width column analysis

Given the long equilibration times required, to further probe the stability of the two double-width structures suggested in the literature, pre-assembled columns were created to match the two dye arrangements proposed. These systems were then solvated, energy minimised using the steepest descent algorithm, and then the natural evolution of the system observed. Referring to figure 4.2, system C corresponds to (a) where the central carbonyl oxygens are facing towards each other, and system D corresponds to (b) where the central carbonyl oxygens are facing away from each other.

System C

System C is visually observed to move from one double-width column to rearrange into two single-width columns (Fig 4.19). The resulting layer structure meets over periodic boundary conditions to form a continuous sheet (via the sulphonate groups), as can be shown in figure 4.20.

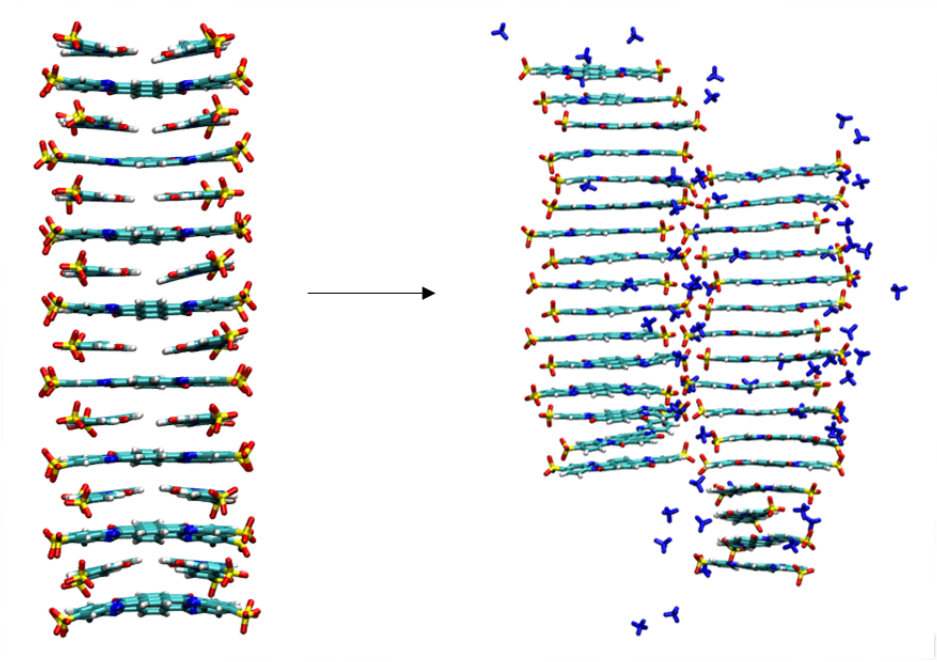

Figure 4.19: System C starting configuration (left) and equilibrated layer structure (right). Solvent has been removed for visual quality, and counterions removed from the starting configuration (left).

Intermolecular distance

The intermolecular distance is the average distance between neighbouring dye molecules in a stack. Analysis of the single width columns formed for system C shows an intermolecular distance of 0.33 nm (Fig 4.21), a value typical for chromonic systems. The isolated columns in solution discussed earlier in the chapter (system A and B), showed a slightly larger intermolecular distance of 0.34 nm, indicating that a Bordeaux dye system, where columns interact with adjacent columns, leads to a small reduction in the stacking distance between neighbours.

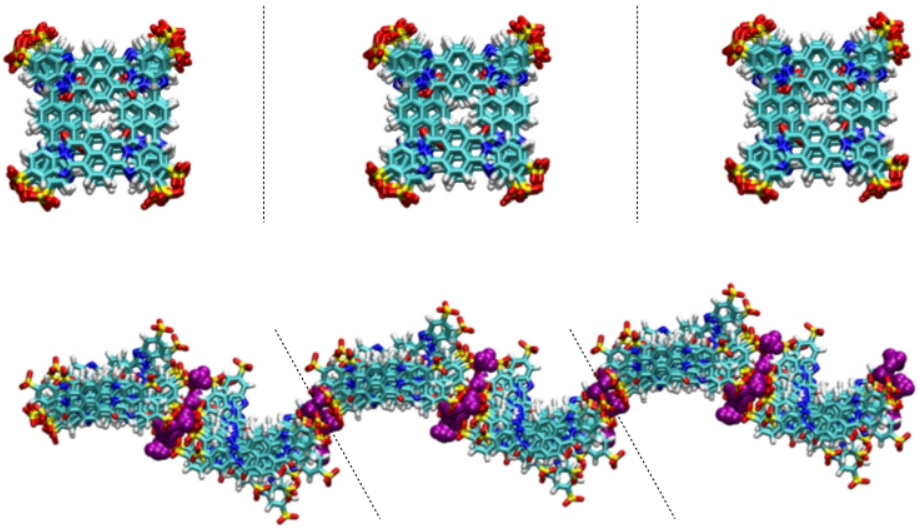

Figure 4.20: Sky view of the initial structure for system C (top) and the final single width structure (bottom). Three simulation boxes are shown (main simulation cell + one either side in the z-direction), with dotted lines used to show where periodic boundary conditions lie. Selected bridging counterions are shown in purple.

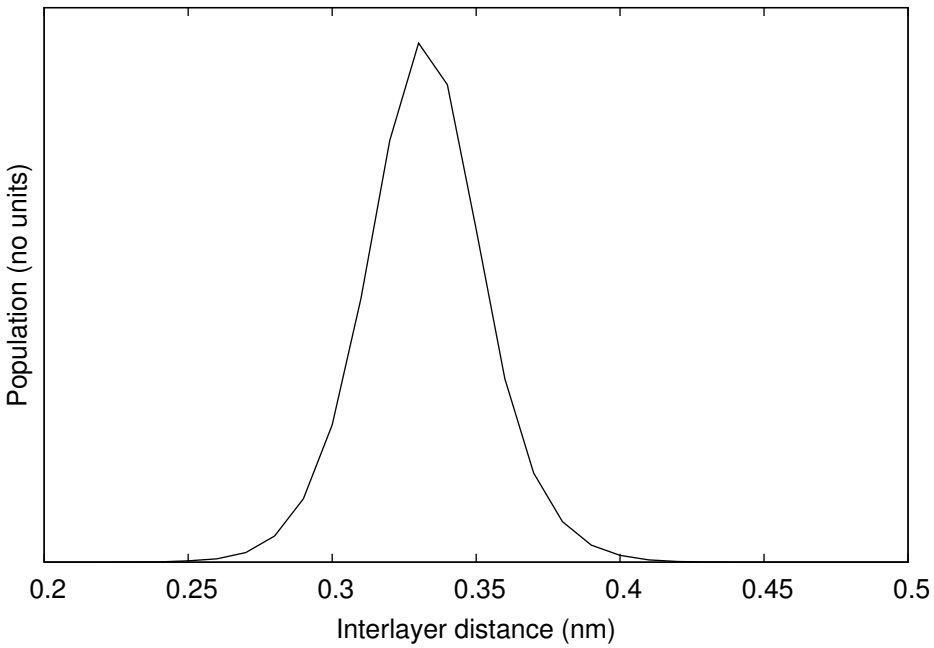

Figure 4.21: Intermolecular distance for system C.

Analysis of twist angle

Analysis of system C, calculating a histogram of population as a function of twisting angle between adjacent and next-adjacent molecules was used to observe, on average, how much the molecules twist with respect to neighbours and next-neighbours (Fig 4.22).

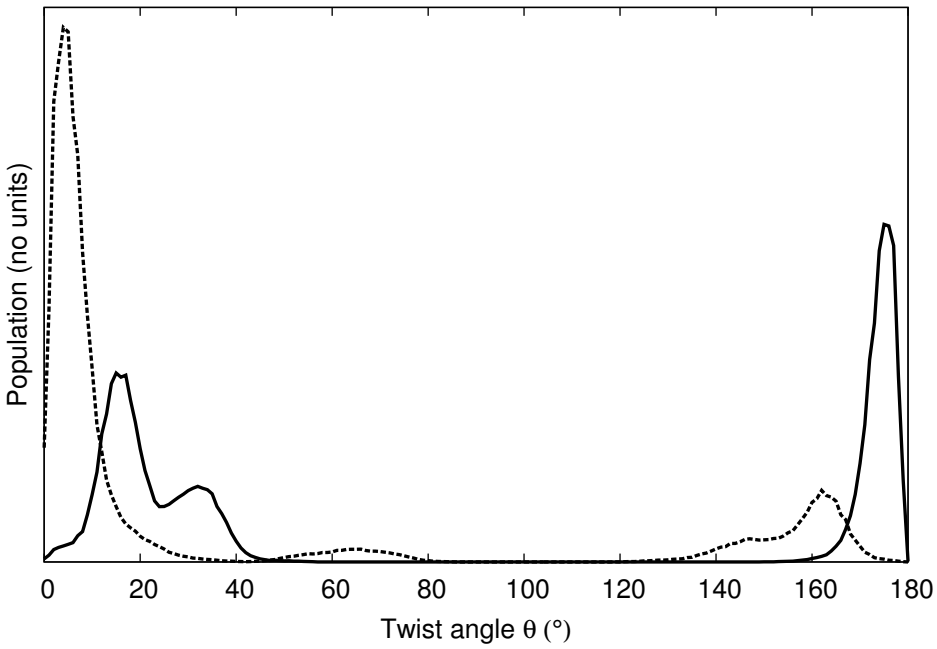

Figure 4.22: Twist angle analysis for system C, with neighbours (solid) and next-neighbours (dashed).

Neighbouring dye molecules show two primary twist regions, with a large peak at 174-177° and an additional bimodal peak at 14-17° and 30-34°. This indicates that upon unravelling from the pre-assembled structure, neighbouring dye molecules adopt both an antiparallel and offset parallel arrangement. The bimodal peak also correlates to the previous findings for system A and system B, whereby neighbours in a parallel configuration can dynamically switch between two preferential configurations. Next-neighbours show a large predominant peak in the 3-6° region, and a smaller peak in the 161-164° region.

Column-column analysis

Analysis of the distance between column centres within system C (fig 4.23), using the method outlined in chapter 2, shows an average spacing of 2.09 nm. This value corresponds to the molecular length along the long axis of a single dye molecule (see fig 4.18).

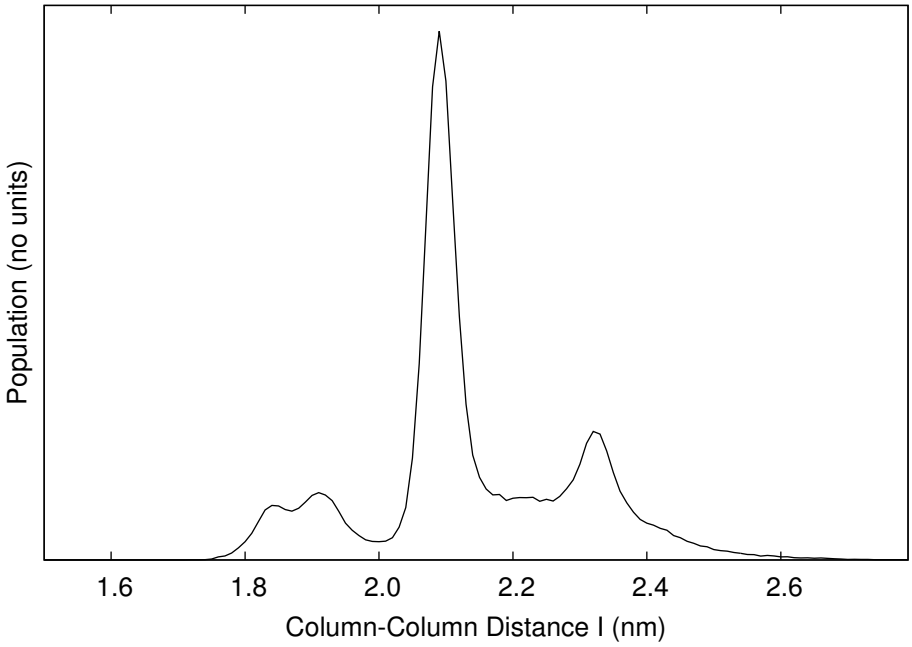

Figure 4.23: Column-column distance for system C.

System D

System D, whilst at first glance appearing to retain its initial double-width column structure over a time period of several hundred nanoseconds, is not a completely stable structure (Fig 4.24). Both ends of the double-width column can be visually seen to have fragmented into two single-width columns, whilst the central region of the seeded structure retains a double-width structure. The slow transition from the seeded structure to single width columns could indicate that both structures are energy minima, but a single width column is lower in energy, and thus transitions to this state over time. The slower nature of column disintegration compared to system C, reveals that of the two possible double-width column configurations, system D is

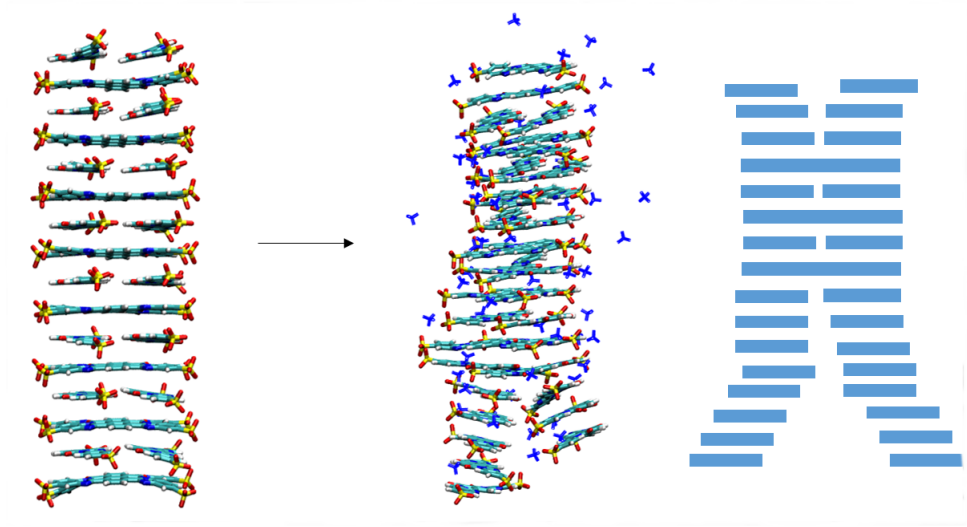

Figure 4.24: System D starting configuration (left) and system evolution after several hundred nanoseconds (middle). A simple schematic is included (right) to help visualise the final structure, and the slow transition to two single width columns occuring.

an energetically more favourable starting structure than system C. Although neither are seen to form spontaneously in any simulation.

Conclusion

Double width columns have been pre-assembled and allowed to evolve in simulations to observe the stability of these structures. The initial structure for system C disassembles fairly quickly over 100 ns to form two single width columns, that interact via ends to form a continuous layer across the periodic boundary conditions. Two repeat distances are recorded, with a stacking distance along the stack of 0.33 nm, and a column-column distance of 2.09 nm. System D shows the long ends of the double width structure to be slowly separating into two separate stacks. The time frame for this process is longer than the many hundreds of nanoseconds already performed, indicating that system D is likely to be more energetically viable than system C, but both are less likely to be as stable as a single column structure. Simulations have shown no evidence for either of the suggested double width column structures for the Bordeaux dye in the literature.

4.3.4 Counterion effect

Radial distribution functions

Radial distribution functions, $g(r)$, were produced for system A, system B and system C to describe the average structure between the SO_3^- group of the dye molecules and nitrogen atom of the counterions (figure 4.25).

The behaviour of the $g(r)$ remains qualitatively similar for all three systems. A high degree of order is observed between the dye molecules and counterions, confirming previous visual evidence of partial condensation of ammonium ions around the aggregate. Radial distributions between the ammonium nitrogen and sulphonate oxygens of the dye show a preferential "binding site" at a distance of 0.27 nm and a secondary site at 0.5 nm. The main peak corresponds to a strong favourable interaction between the ammonium nitrogen atom and the three oxygens of the dye sulphonate group. The second peak corresponds to the counterion interacting with the sulphonate group on one of the dyes either side of the reference dye. The primary site for the radial distribution of the ammonium nitrogen atom and the sulphur of the sulphonate group is 0.41 nm. system A and system B show a secondary site at 0.6 nm, and system C shows a secondary site at 0.75 nm.

Bridging and alteration of the counterion

As observed in the system C, the counterions appear to form a bridging network between the two columns in solution, via the periphery sulphonate groups of the dye molecules (fig 4.19). The bridging of highly charged rods via a counterion is typically observed in DNA, where multivalent-ion mediated attraction between similarly charged DNA molecules can occur.[102–109] The two tentative models suggested for this mechanism are transient bridging, when a multivalent counterion bridges two neighbouring rods,[106] and the formation of sticky regions where counterions are condensed at the rods.[103,105,107,110,111] For ionic chromonic systems, the presence of spermine salt $SpmCl_4$ to a Sunset yellow solution results in Spm^{4+} acting as inter-rod linkers, bridging the neighbouring SSY aggregates in solution.[6] Under specific conditions, the linker-assistant aggregates can even form filaments and bundles.[6,112]

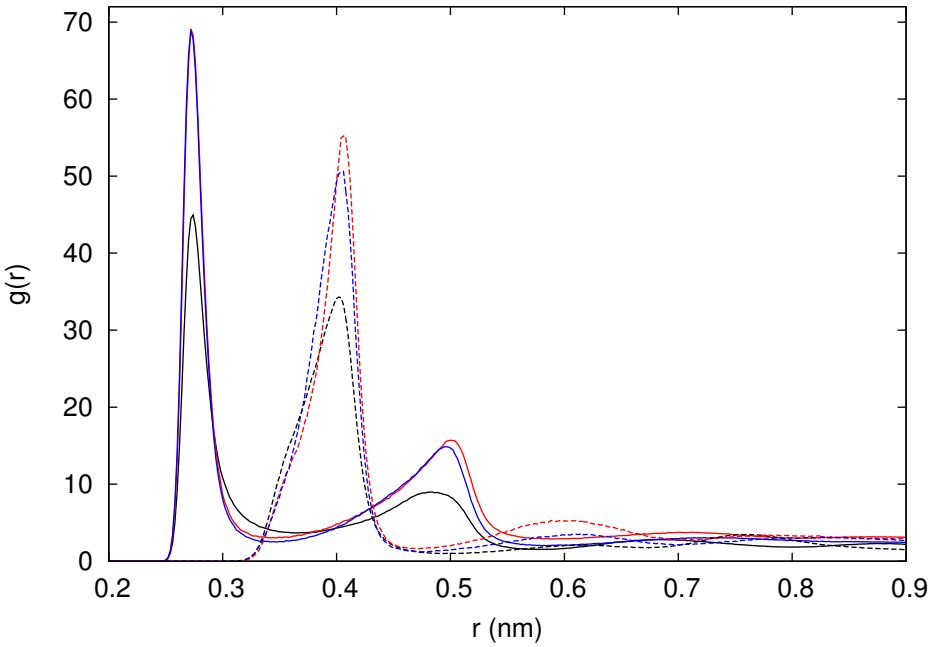

Figure 4.25: Radial distribution functions for system A (red), system B (blue), and system C (black) between the counterion N and dye oxygens (solid lines) and counterion N and dye sulphurs (dashed).

In order to examine the bridging strength, the final structure from system C was removed and placed into a larger simulation cell, and solvated to produce a 10.9 wt.% dye concentration. The trajectory showed the columns immediately moving apart in solution. Concluding that the counterions are not sufficiently strongly binding to stabilise a double column structure in more dilute solution, it should be also noted that unless several columns are aligned next to each other and forming a more fixed network, such as the previous layer formed with periodic boundary conditions causing a continuous layer, it is entropically favourable to have the columns freely move in solution.

Two further seeded double-width columns systems were created using the same preassembled structures as systems C and D (named system E and system F) but with the ammonium counterion replaced with ethyl ammonium, a larger counterion with the inability to hydrogen bond. Evolution of the system shows both ends of the stacks moving from a double-width structure to a single-width structure (fig 4.26). Unfortunately, due to the simulation cell size, upon disintegration of the preassembled structure the ends of columns meet via periodic boundary conditions

to fix the structure in place. This results in a structure of predominant single-width arrangement, interspersed with frozen regions of double-width arrangements. The counterions are not seen to associate or group around the charged sulphonate region on the dye molecules. No bridging is observed, due to the system not forming two single columns as a result of being trapped in the structure and not reaching equilibration.

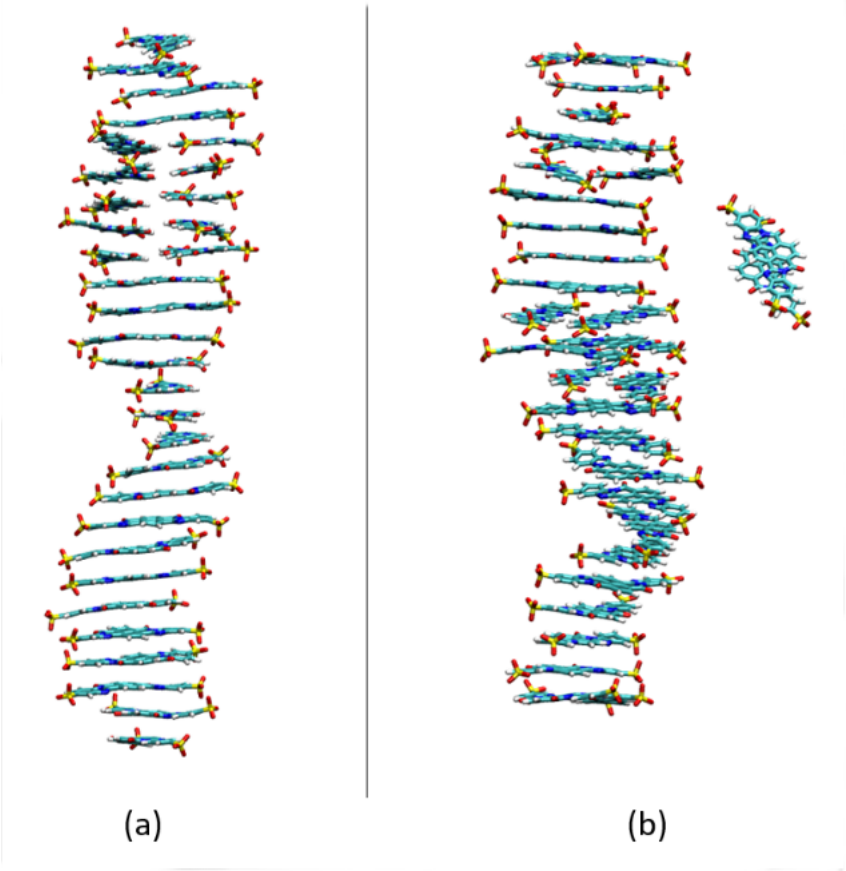

Figure 4.26: System E and system F with alternative counterions ethyl ammonium. Dye molecules shown only for visual clarity.

4.3.5 Conclusions

Extensive atomistic level molecular dynamics simulations, for the Bordeaux dye in aqueous solution, have been used to provide a molecular level picture of the preferred stacking structure within chromonic aggregates. The dye molecules have been found to self-assemble to form aggregates in solution. The dyes come together in a face-to-face fashion, with a stacking distance of 0.34 nm between neighbouring molecules.

A predominantly antiparallel arrangement of 174-177° between neighbouring dye molecules is observed, allowing the charged sulphonate groups to minimise repulsion caused by close contact. However, dynamic parallel configurations, with switching between two stable structures corresponding to 10-15° and 30-35° angles, are also observed. This indicates that the binding between dye molecules is much stronger than the repulsion caused by close contact sulphonate groups.

A thermodynamic study of the Bordeaux dye showed an extremely strong binding energy between dye molecules of 27.1 $k_{\mathrm{B}}\mathrm{T}$, a value significantly larger than typically found for chromonic systems (7-14 $k_{\mathrm{B}}\mathrm{T}$). The binding energy for Bordeaux dye in the literature,[5] is commensurate with the computational value corresponding to the energy change from orienting from a H-aggregate (global minimum) to a J-aggregate (local minima). The presence of both H- and J-aggregate arrangements indicates that the Bordeaux dye has a much more complex model of aggregation behaviour than the simple H-aggregate assumed for many of the exprimentally derived thermodynamic values.

Two seeded double width column structures were used to test the integrity and feasibility of such a structure in solution. System C transitioned to a double column layer structure, with counterions potentially acting as linkers between the two columns of aggregates. System D demonstrated more structural integrity than system C, with many double width structural units being retained for the entirety of the simulation, suggesting a energetically favourable local minima structure. No double width column structure was observed to occur spontaneously from monomer aggregation, suggesting that the single width column is the most energetically favourable aggregation mode.

Chapter 5

Coarse graining of chromonic systems

5.1 Background to coarse graining

5.1.1 Computational accessibility of chromonic liquid crystal phenomena

In chromonic liquid crystals, a vast range of phenomena occur over many different time and length-scales, from the aggregation of a few dye molecules in solution, to the movement of many thousands of columns to form an ordered phase. A balance must be struck between the level of accuracy required to simulate a particular phenomenon and its associated computational cost. In particular the time and length-scale achievable depends on the complexity of the model used (as shown in fig 5.1). A quantum mechanical approach, such as density functional theory, allows for electronic structure to be described and the spread of charge across a single chromonic mesogen to be studied. A high level of accuracy is achieved for the energy minimised molecular structure and electronic properties. However, the system is restricted to a small size (typically a few molecules at most). A classical mechanics approach, whereby spheres (atoms) are represented by connected springs (bonds), allows for a larger system size of potentially hundreds of chromonic mesogens solvated by many thousands of water molecules. The level of accuracy is now reduced in comparison

to a quantum approach, since the electrons are treated implicitly. However, the system can now explore the self assembly process and the aggregation mode between molecules within a stack. Accessing large scale phenomena, such as phase transitions, requires hundreds if not thousands of columns interacting in solution. This is far beyond the reach of current atomistic modelling, thus introducing the need for a coarse grained (CG) model.

An example CG model is the recent work on a non-ionic chromonic liquid crystal system, TP6EO2M, using dissipative particle dynamics (DPD) to provide the phase behaviour across a full range of concentrations.[113] DPD simulations consists of particles interacting according to simplified force laws, composed of soft repulsions, pairwise dissipation forces and matching pairwise random forces.[114] Whilst this technique allows for very large systems to be explored, a more systematic approach of using experimental data, or a more detailed reference system, to build a coarse grained model allows for a system-specific potential to be developed. This is an area yet to be examined.

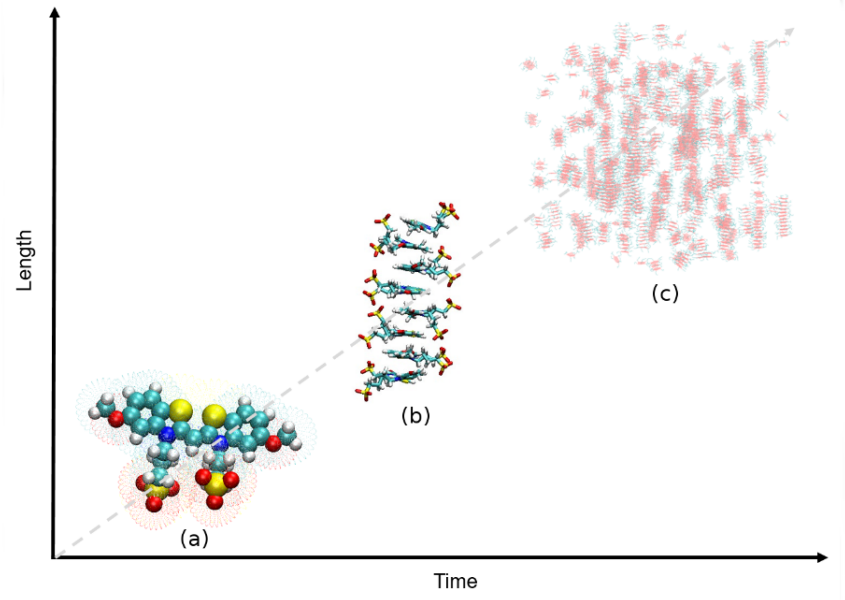

Figure 5.1: Length vs time graph to demonstrate how system size and associated computational cost determine the level of detail accessible in a simulation of a chromonic liquid crystal system: (a) quantum mechanics regime, (b) atomistic simulation regime and (c) coarse-grained simulation regime.

5.1.2 Coarse-grained models

Coarse graining is a technique by which a system can be represented by a reduced number of degrees of freedom, in comparison to an all-atom description. Using this technique, many atomistic sites can be embedded into a single interaction site. The elimination of finer interaction details can provide a significant reduction in the computational cost, whilst retaining the essential physics of the system. Two major computational savings arise. Firstly, the reduced number of interaction sites in the system results in less force calculations at each time step. Secondly, a coarse-grained model often results in a simpler potential landscape, allowing for a shorter time period for a system to reach equilibration. As a result, systems can be simulated at time and length-scales that are orders of magnitude larger than in an atomistic model. The two commonly used approaches to constructing a coarse-grained model are known as either a *top-down* or a *bottom-up* approach.

5.1.3 Coarse-grained mapping

Mapping an atomistic to a coarse-grained representation is an important choice heavily determining whether the model will be successful. A delicate balance must be struck between speeding up the simulation by reducing degrees of freedom, and retaining the essential physics of the system. The behaviour of complex molecules, such as proteins[115,116] and nucleic acids,[117,118] can often be governed by lower level degrees of freedom, making detailed system knowledge necessary for a successful mapping scheme, and thus ultimately a successful coarse-grained model.

The spectrum of mapping choices ranges from an all-atom approach, whereby parameters for every type of atom in the system are included, to the most extreme form of coarse-grained models with whole molecules being encapsulated into a single bead. One of the simplest mapping schemes commonly implemented is the united atom (UA) scheme, in which aliphatic (non-polar) carbon atoms and their hydrogen atoms are treated as a single pseudo-atom. In view of how abundant hydrogen atoms are, particularly in organic molecules, a united atom is a simple alteration to an all-atom approach, which significantly reduces the degrees of freedom and

increases simulation speed. A selection of united atom force fields currently used are the OPLS-UA,[119] Amber united-atom,[120] and all of the GROMOS force fields.[121]

On the far end of the possible mapping schemes are studies where extremely large coarse grain beads are used, typically termed "ultra-coarse-graining" (UCG). For biological systems, this commonly refers to where more than one amino acid is reduced to a single interaction site. However, work by Dama *et al.* has recently proposed a systematic variational UCG method that coarse grain entire protein domains and subdomains (encompassing hundreds of heavy atoms) into a single bead.[122] For complex molecules, such as chromonics, a mapping scheme, which is inter-mediate between these two extremes, is most suitable. Detailed knowledge of the molecule and molecular behaviour allows varying sized groups of atoms to be coarse-grained, in an attempt to capture the essential physics of that particular molecular interaction.

5.1.4 *Top-down* methods

A *top-down* approach is where experimental thermodynamic data is used to guide the parameterisation of the coarse-grained potentials. The two most prominent *top-down* approaches are the MARTINI force field[123] and methods based on fitting experimental data. The latter includes the use of liquid state theories such as the Statistical Associating Fluid Theory (SAFT).[124]

The MARTINI force field, developed by Marrink *et al*, consists of pre-defined interaction sites which are based on reproducing the partitioning free energies between polar and apolar phases of a large number of chemical compounds.[123] On average, 4 atoms are combined into a bead, with each bead being assigned one of four basic interaction descriptions: polar (P), non polar (N), apolar (C) and charged (Q). The description is then extended to include terms introducing chemical traits such as hydrogen bonding capability or degree of polarity. An example mapping and interaction description scheme can be seen for a selection of molecules in fig 5.2. The MARTINI force field has been implemented frequently for coarse graining biological systems, such as carbohydrates,[125] proteins[126] and collagen molecules.[127] Key problems are best exemplified by the behaviour of water, which commonly freezes upon

the surface of the solute. The use of a Lennard-Jones (LJ) 12-6 potential typically creates potentials sharper than the potentials needed for a coarse-grained model, which are inherently softer than those typically used for atomistic potentials.

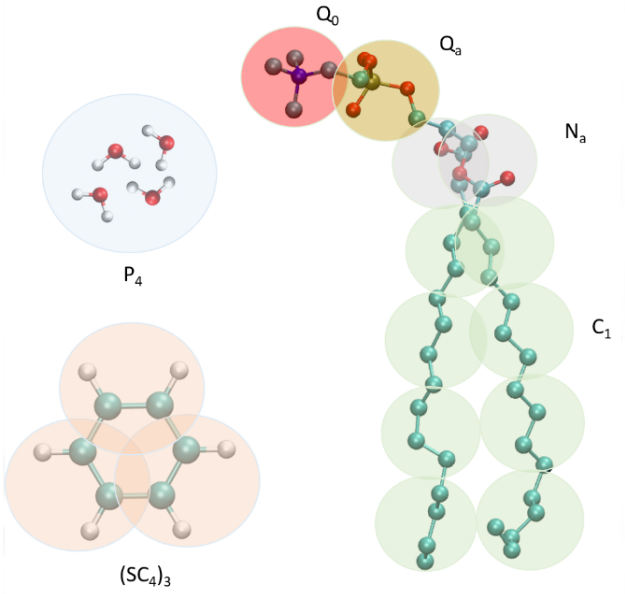

Figure 5.2: A MARTINI representation for coarse-grained water (top left), benzene (bottom left) and a DPPC lipid (right).

SAFT is an equation of state, where the essence of the theory is that the Helmholtz free energy is a summation of expressions encompassing not only the effects of repulsion and dispersion forces, but also association and solvation effects[124] (as summarised in the schematic in fig 5.3). Fitting model parameters to theory allows for the parameterisation of coarse-grained potentials.

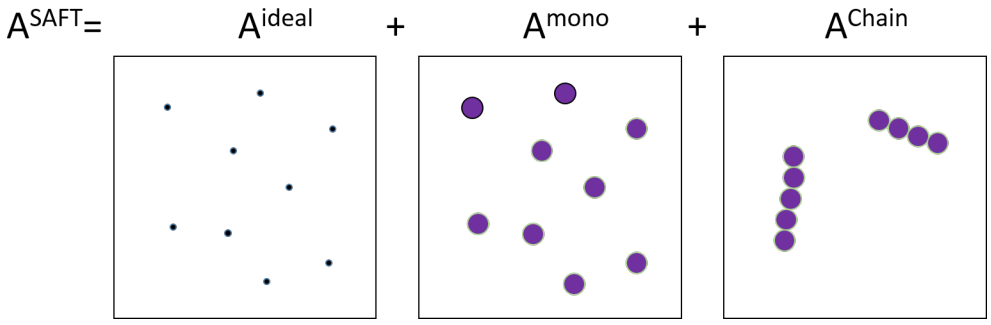

Figure 5.3: Schematic diagram to show the contributions to the overall Helmholtz free energy in SAFT.

5.1.5 *Bottom-up* methods

A *bottom-up* approach is where the coarse-grained potentials are developed by reducing the number of degrees of freedom of a higher level reference system (an atomistic system). The most typical *bottom-up* methods parameterise a potential to reproduce key thermodynamic or structural features of the reference system.

Structure-based methods

Boltzmann Inversion

Boltzmann inversion is the simplest structure-based approach available to obtain coarse-grained potentials.[128] The method arises from the fact that independent degrees of freedom q obey the Boltzmann distribution in a canonical ensemble, where $P(q)$ is the probability measure of the degree of freedom, q.

$$P(q) = Z^{-1} \exp[-\beta U(q)], \tag{5.1}$$

where $Z = \int \exp[-\beta U(q)] dq$ is a partition function, $\beta = 1/k_B T$. Inversion of equation 5.1 gives rise to a potential of mean force, $U(q)$:

$$U(q) = -k_B T \ln P(q). \tag{5.2}$$

In practice, $P(q)$ is typically a CG bond, angle or dihedral distribution or a non-bonded radial distribution function (RDF). A main assumption, and the reason why Boltzmann inversion is ordinarily only used for bonded interactions, is that the interactions in the system are not correlated and can be treated separately.[129]

Iterative Boltzmann Inversion

The weakness of Boltzmann inversion is being able to treat correlated potentials, which can be improved by introducing an iterative scheme for the non-bonded potentials,[130]

$$U^{n+1} = U^{(n)} + k_B T \ln \frac{P^{(n)}}{P_{\text{ref}}}. \tag{5.3}$$

The iterative Boltzmann inversion (IBI) approach can be implemented in a series of steps. Firstly, an initial guess for the potential P^n is made by Boltzmann inversion. Secondly, a short coarse-grained simulation is carried out with this potential, and the potential is then updated with respect to the difference between the current RDF and the target RDF (P_{ref}). This process is repeated until the RDF for the coarse-grained simulation sufficiently matches the reference RDF. An example of the iteration process can be seen in figure 5.4.

The main weaknesses of IBI are the non-transferability of the potentials to systems of different concentrations than the reference and convergence issues when dealing with a large number of different interactions in a system.

Incorporated into the potential update function is also a term to correct the pressure of the system, using the following linear correction function:

$$\triangle U(r) = A\left(1 - \frac{r}{r_{\text{cut}}}\right), \tag{5.4}$$

$$A = \text{sgn}(\triangle P)0.1k_BT\min(1, \mid f\triangle P). \tag{5.5}$$

The goal of IBI is to reproduce the distribution functions and therefore structure of the reference system. It succeeds in providing a method to deal with correlated potentials. However, too many interaction sites in the coarse-grained system can lead to convergence issues. Improving a particular interaction potential could unfavourably modify the RDF for a different interaction, with this risk increasing heavily as the number of interactions for the coarse-grained system increases. Hence, finding optimum potentials can be compared to the problem of finding a global minimum energy structure in a complex potential energy landscape.

Force matching

Multi-scale Coarse Graining (MS-CG)

Multi-scale coarse graining (MS-CG), often referred to as force matching (FM), is a non-structured based and non-iterative approach to developing a coarse grained potential from an atomistic reference system. The aim of force matching is to

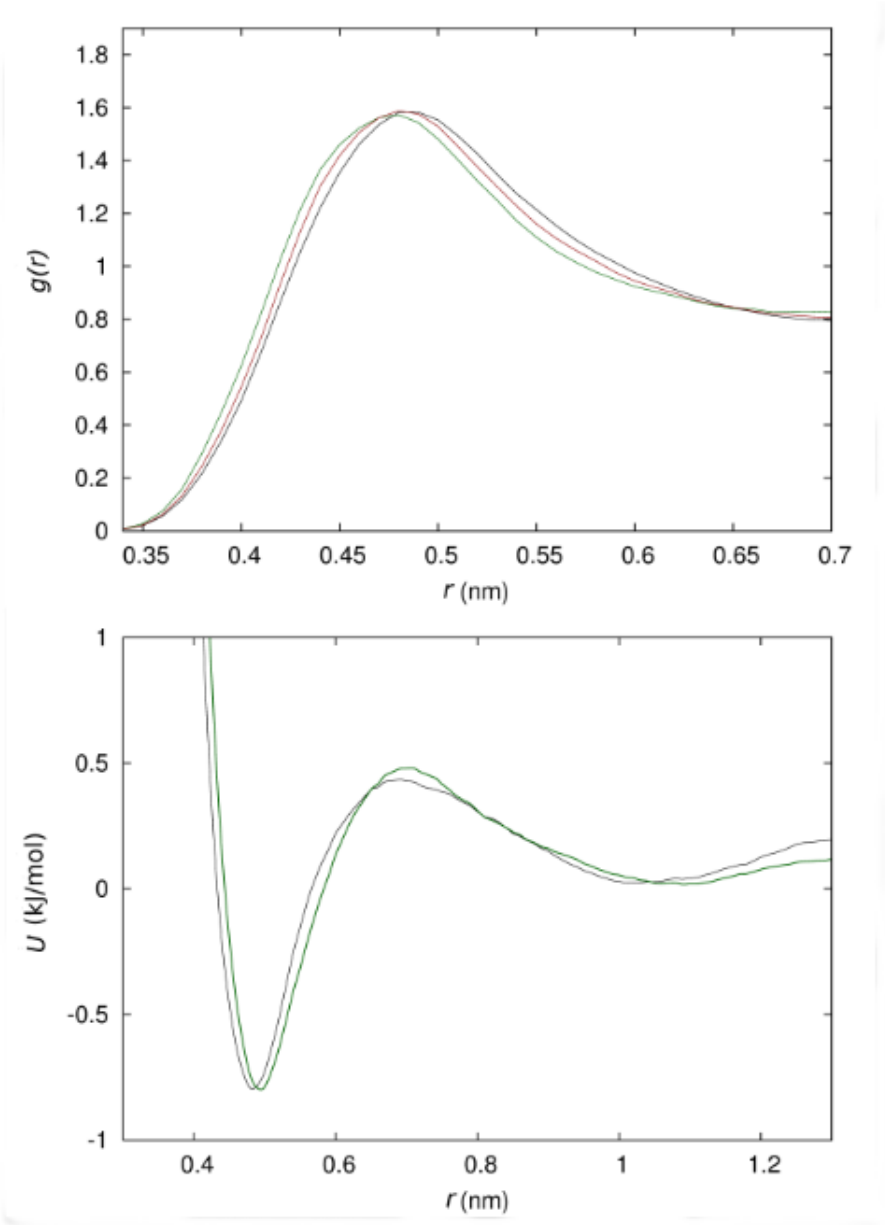

Figure 5.4: Example of the iterative Boltzmann inversion process using a single non-bonded interaction within a pure hexane system. The top graph shows the target radial distribution function (red) and two RDFs as example iterations (black and green). The bottom graph shows the potential for the two selected iterations.

evaluate the inter-atomic forces from an atomistic simulation, and map this onto the corresponding coarse grained interaction sites.[131–133]

The method works by first assuming that the coarse-grained force field, and therefore the forces arising from it, depend on M different parameters $g_1 ... g_M$. These parameters are coefficients of splines used to describe the functional form of the forces. Determining these parameters can be achieved by properly re-weighting the forces from the reference system

$$\mathbf{f}_i^{\text{ref}} = M_i \sum_\alpha \frac{w_\alpha \mathbf{f}_\alpha}{m_\alpha}, \tag{5.6}$$

$$M_i = \sum_\alpha \frac{m_\alpha}{w_\alpha^2}, \tag{5.7}$$

where M_i is the mass of the bead i, index α numbers all atoms in this bead, $\mathbf{f}_\alpha$ is the force on atom α, m_α is corresponding mass of that atom, w_α are mapping coefficients used to obtain the position of the coarse-grained bead $\mathbf{R}_i$, where

$$\mathbf{R}_i = \sum_\alpha w_\alpha \mathbf{r}_\alpha. \tag{5.8}$$

Using the centre of mass of each coarse-grained mapped site, the reference forces can simply be equated to the sum of the atomistic forces. Calculating the reference forces for a large number of snapshots allows the coarse-grained potential to be generated.

5.1.6 Aims of this chapter

The overall aim of this chapter is to attempt to coarse grain chromonic mesogen Dye A (chapter 3). In view of the wealth of atomistic simulations performed on Dye A in chapter 3, with results giving good comparison to experimental data, a *bottom-up* approach will be used to generate coarse grain potentials.

A comparison will be made between the force matching (FM) and iterative Boltzmann inversion (IBI) procedure applied, with aims to determine which of these two procedures are perhaps more suitable for a system as complex as an ionic chromonic.

A contrast will also be made between the limitations and problems experienced with both procedures, with suggestions for possible solutions. Taking into account the complexity of coarse-graining such a large molecule, where so many different functional groups and interactions are at play, a variety of mapping schemes will be used. The more refined (less atoms per bead) the mapping system for the dye is, the more interactions present in the system. The balance between the mapping resolution and the ability to represent overall system behaviour will be probed, with the aim of mapping the chromonic dye in the most efficient way, while maintaining the aggregation and intra-columnar behaviour observed in the atomistic reference systems.

Two simpler homogeneous systems of liquid hexane and water will first be studied using both FM and IBI, with progression onto hexane/water mixtures. The study of these mixtures will allow for a less complex, phase separated system in water to be explored, before advancing onto chromonic systems. This chapter hopes to provide answers as to firstly, whether an ionic chromonic liquid crystal can be successfully coarse grained using the methods discussed, and if so, what steps and mapping schemes are needed to create a successful coarse grained set of potentials.

5.2 Computational details

For the atomistic simulations, the potential functions and interaction parameters were employed by the GAFF[60,80] parameters (described in section 3.2) as implemented by the simulation suite Gromacs.[81–85] The partial atomic charges were calculated from BCC partitioning with the AMBER sqm (semiemperical quantum mechanics) programme, which carries out single point calculations and energy minimisations (geometry optimisations).

Molecular dynamics simulations for the atomistic systems were initially equilibrated in a canonical (NVT) ensemble and then followed by an extensive equilibration run using the isobaric-isothermal (NPT) ensemble. All simulations were performed at a temperature of 300 K, with a Nosé-Hoover thermostat[65] being employed. Pressure was controlled with a Parrinello-Rahman barostat[69] at atmospheric pressure with an isotropic pressure coupling. A time step of 1 fs was used for all atomistic simulations. The interaction cut-off for Lennard-Jones interactions was 1.1 nm, while short range Coulombic interactions were truncated at 1.2 nm. The long range part of the electrostatic interactions was represented by employing a Particle Mesh Ewald (PME) summation.[55] The TIP3P[90] water model was chosen as most suitable for this system, due to accurate aqueous membrane bilayer simulations demonstrating that the GAFF/TIP3P combination could be utilized in biomolecular modelling.[91]

Initial simulations consisted of a system of 256 hexane molecules randomly oriented and positioned in a cubic box, and a system with 1821 water molecules in a cubic box. Three hexane in water systems were simulated, where the smallest system (HS1) contained 50 hexane molecules, the middle size (HS2) had 200 hexane molecules, and the largest system (HS3) had 400 hexane molecules. Each system had the same hexane concentration of 11.6 wt.%. Each system was equilibrated for tens of nanoseconds, before a production run of several nanoseconds was then performed. The chromonic Dye A system contained a stack of 8 dye molecules in solution, at 9 wt.%. This system was run for many hundreds of nanoseconds (as shown in chapter 3), and a production run of 1 ns performed for the data to be used

in the coarse graining methods, where frames with both structural and force data were outputted every 100 fs.

Coarse-grained potentials were developed using the VOTCA package.[129] For the IBI technique, each interation had a minimum of a 0.1 ns simulation, with larger systems such as HS3 having up to 0.5 ns simulation time per iteration. While the IBI failed to converge all potentials to the target radial distribution functions, the iteration with the best overall convergence criteria match has been selected for all three systems. After a minimum of 200 interations, iteration 238 was selected for HS1, iteration 198 for HS2 and iteration 63 for HS3. After a minimum of 50 iterations for dye systems, iteration 33 was selected for dye system M1, and iteration 7 for dye system M2. Further iteration steps from these selected iterations showed no more improvement in converging the potentials, with large fluctuations in convergence match criteria occuring for each iteration. CG simulations with the user-defined potentials were all run in a canonical (NVT) ensemble, with all other simulation parameters matching the atomistic simulations described above. Equilibration was reached within several nanoseconds, and a production run of 1 ns completed for analysis purposes. For the force matching technique, forces were averaged over 20 frames per block, with a minimum of 400 force calculations done until convergence was observed between blocks.

Various analysis methods were used to quantitatively assess the structure of the chromonic aggregates. A detailed explanation of these various analytical techniques can be found in the end of Chapter 2. The unit vector defining the normal to the plane of the molecules, was defined by atoms N, S and C2 (highlighted in red in Chapter 2, figure 1.5) for the atomistic system, and interactions sites D1, E and D2 for the M2 CG system. The unit direction vector (for the angle calculation) is defined by atoms C3 and C6 (highlighted in blue in Chapter 2, figure 1.5) for the atomistic system, and interaction sites D1 and D2 for the M2 CG system. To achieve a stable centre of mass, the flexible arms were removed for the analysis.

5.3 Results and discussions

5.3.1 Homogeneous systems

Before attempting to coarse grain complex chromonic systems, a number of simpler systems were selected to study first. Exploring simpler systems allowed increased levels of complexity to be added in manageable stages, with software issues being more easily resolved when focusing on less complex systems. The simplest systems studied here are a pure hexane and a pure water system, followed by hexane/water mixtures where phase separation has occured. It should be noted that pure hexane and pure water systems have been studied in the literature previously using the VOTCA package,[129,134] but will be recalculated here for comparison purposes and to provide a basis for tackling more complex systems. The force matching technique applied here is a hybrid-FM technique, where the bonded potentials are generated from Boltzmann inversion, and the non-bonded potentials from force matching. Previous work has shown that a hybrid approach for force matching, where only the intermolecular contribution to forces were considered, achieved the best results.[134]

Hexane system

The structure and mapping scheme for hexane can be seen in Figure 5.5. A 3-site representation was used, with each interaction site containing 2 heavy atoms and its associated hydrogen atoms, giving rise to two symmetrical outer beads (named A), and an inner bead (named B). This representation provides 1 bond (AB) and 1 angle (ABA), of which the potentials can be seen in Figure 5.6. Boltzmann inversion from an atomistic hexane simulation was used to generate the bonded potentials. The 3 non-bonded potentials for the A-A, B-B and A-B interactions were generated using both IBI and FM.

Results for the coarse grain hexane simulations show that both IBI and FM structurally reproduce the atomistic reference system, as can be shown by the radial distribution functions (RDF) in figure 5.7. Both coarse graining techniques have produced a set of potentials resulting in a coarse grained representation that matched the reference system to a high degree of accuracy. It should be noted that whilst

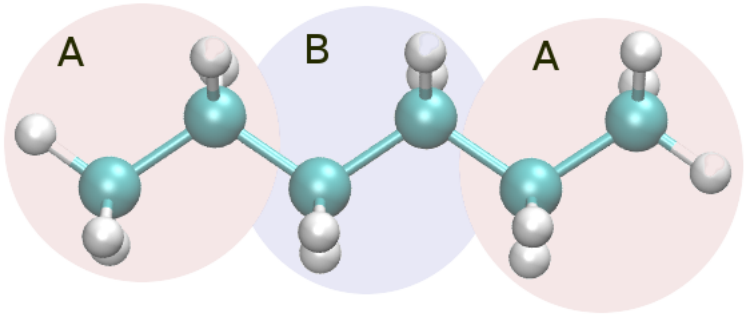

Figure 5.5: Coarse grained mapping scheme for hexane. Each interaction site contains 2 heavy atoms and their associated hydrogens.

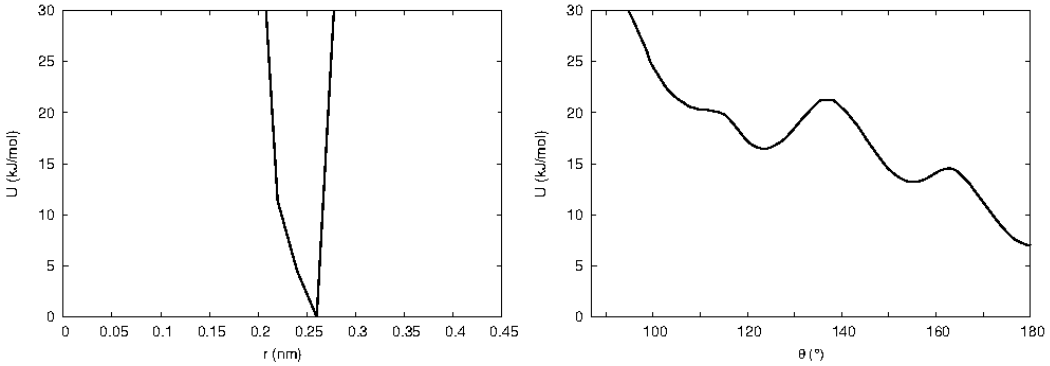

Figure 5.6: Bond (left) and angle potential (right) for coarse grained hexane, generated from Boltmann inversion of an atomistic hexane simulation.

the RDFs are matching the literature, the FM potentials differ from the literature due to the additional pressure correction applied here. This can be observed for the A-A potential in fig 5.8. whereby the pressure correction of the initial potential from force matching seen here and in the literature (black), results in a raised well-depth and tail to lower the overall attraction, thus reducing the pressure.

Water system

The structure and mapping scheme for water is a 1-site representation, with the entire molecule encapsulated into a single bead (named C). This representation provides no inner degrees of freedom, such as bonds and angles, and has only 1 non-bonded interaction present, C-C.

Results for the coarse grained water systems show that IBI and FM differ in their ability to reproduce the structural properties of atomistic water, as shown by the radial distribution functions in figure 5.9. IBI is able to fully reproduce the atomistic target RDF, and whilst FM is able to successfully reproduce the initial solvation shell at 0.28 nm, with a peak of 0.29 nm, the longer range structural properties are not fully captured, with FM introducing a secondary solvation shell at 0.38 nm.

5.3.2 Heterogeneous systems

In the next step of introducing further complexity to the system, a hexane/water mixture that phase separates has been coarse grained using the IBI and FM methods. This is already a more challenging system than those typically used in many previous literature studies, which have tended to concentrate on homogeneous solutions and mixtures. The structure and mapping scheme for hexane and solvent are as in the homogeneous systems, with a 3-site hexane, and a 1-site water representation. The results for both the IBI and FM coarse graining schemes will be discussed separately and compared, as well as a comparison to the homogeneous hexane and water systems discussed earlier. The atomistic reference systems, that the coarse grain system will attempt to structurally reproduce, shows hexane molecules reduc-

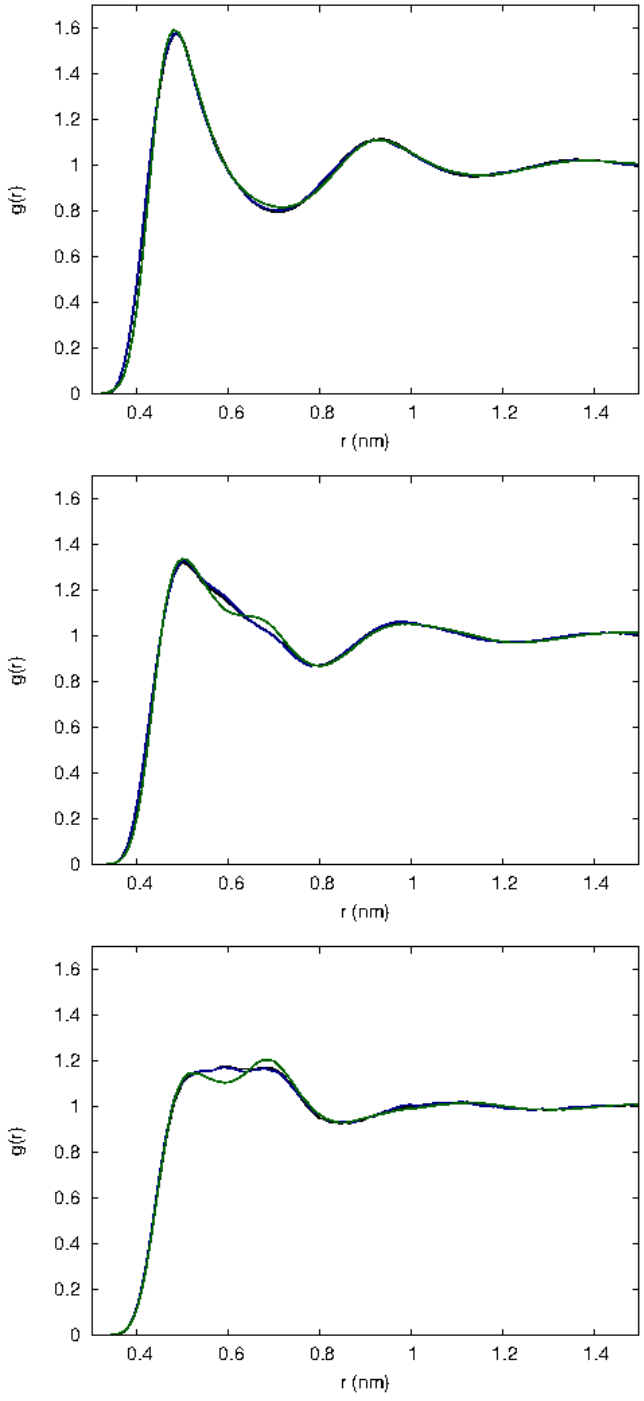

Figure 5.7: Radial distribution functions for the 3 non-bonded interactions within a coarse grained hexane system: A-A (top), A-B (middle) and B-B (bottom) where atomistic (black), IBI (blue) and FM (green) distributions are shown.

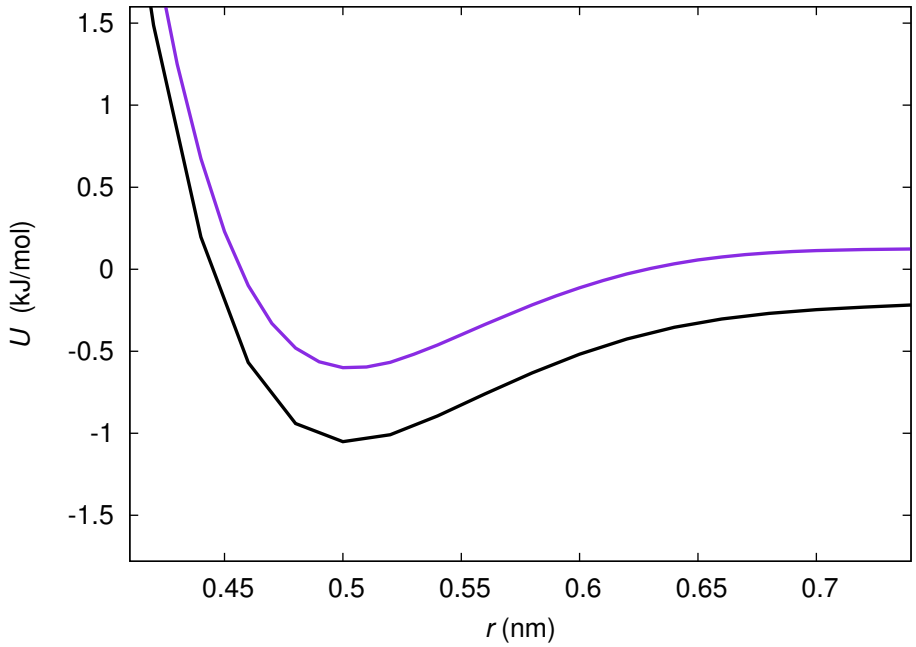

Figure 5.8: Comparison of the A-A potential for hexane from force matching with (purple) and without (black) applied pressure correction

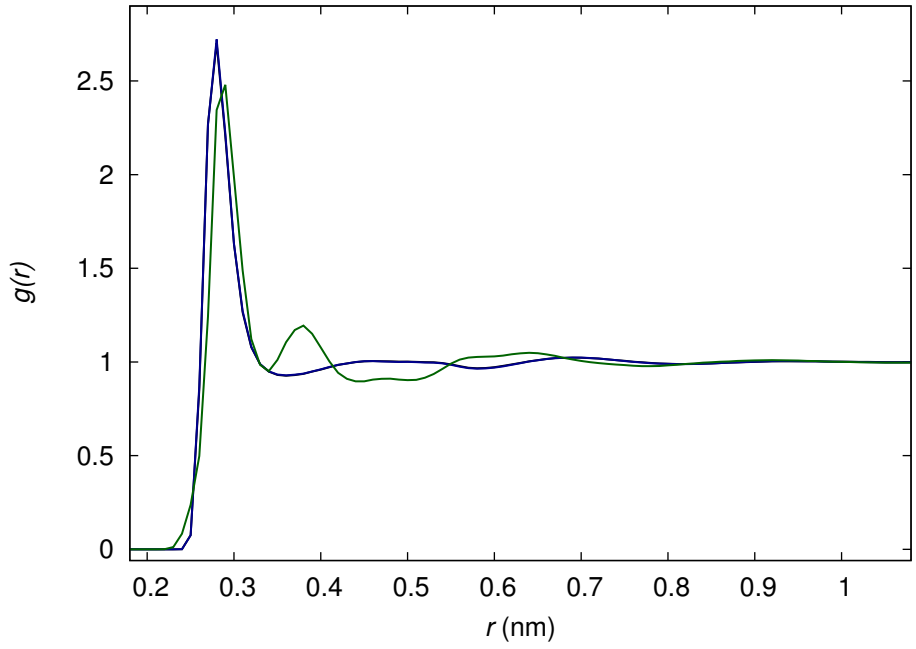

Figure 5.9: Radial distribution functions for the non-bonded interaction (C-C) within a coarse grained water system, where atomistic (black), IBI (blue) and FM (green) distributions are shown. The black and blue lines are perfectly overlapping.

ing contact with the water molecules by phase separating to form a ball of hexane within the water.

Hexane/water mixture using IBI

A system of hexane in water containing 50 hexane molecules (HS1) at an 11.6 wt.%, was coarse grained using IBI to generate coarse grain potentials. While the IBI failed to converge all potentials to the target radial distribution functions, the iteration with the best overall convergence criteria match has been selected for all three systems. After a minimum of 200 interations, iteration 238 was selected for HS1, iteration 198 for HS2 and iteration 63 for HS3. Further iteration steps from these selected iterations showed no more improvement in converging all the potentials.

Visual inspection of the CG simulation shows the hexane phase separating to form channels of hexane branching out into the water (figure 5.10, left). Whilst the CG potentials have reproduced the general phase separation of the two components, it has failed to attain the lowest energy structure that hexane would adopt from the atomistic simulation (a sphere). The system shows poor convergence between the coarse-grained and atomistic RDFs, as shown in figure 5.11. The RDFs for each interaction shows the CG RDFs qualitatively match the shape of the atomistic reference function, while differing in the density of particles at each distance. Hexane-hexane interactions within the system (A-A, A-B, B-B) significantly underestimate the distribution functions, in sharp contrast to the homogeneous hexane system earlier, which perfectly matched for each hexane-hexane interaction. The water-water interaction (C-C) shows a perfect structural match with the atomistic reference system, as with the pure homogenous water system from earlier. Cross-terms for hexane-water interactions (A-C, B-C) both show an overestimation in the CG RDF compared to the reference. In summary, the CG potentials generated IBI for HS1 have not managed to sufficiently reproduce the structural aspects of the atomistic reference simulation.

The failure to sufficiently reproduce the HS1 atomistic reference system can be attributed to IBI method being system size dependent when dealing with a two-component system that phase separates. The small size of the hexane sphere in HS1

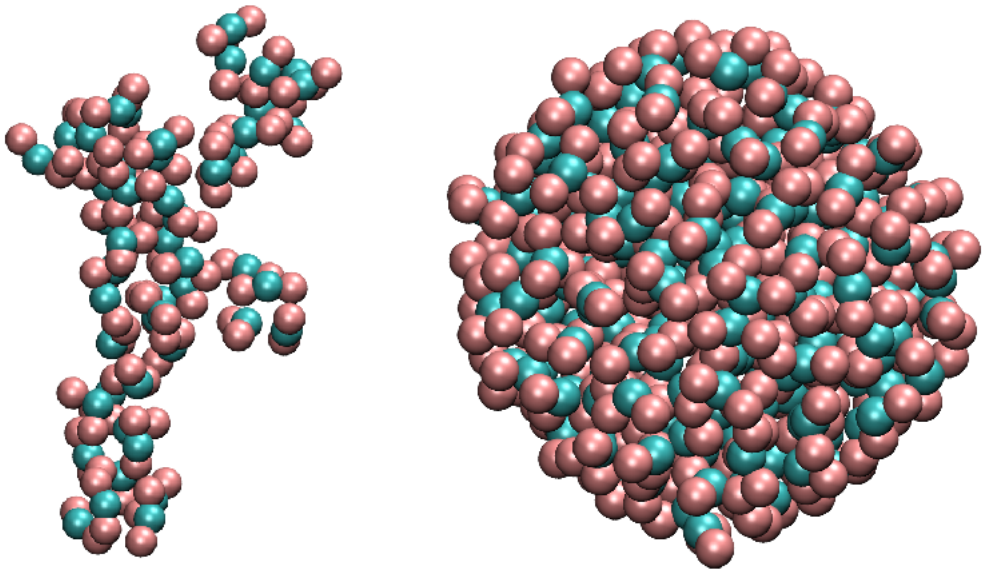

Figure 5.10: Coarse grain hexane structure for HS1 (left) and HS2 (right). Water has been removed for visual clarity.

provides less information on bulk hexane-hexane interactions, and more between the surface of the hexane sphere and the water. This provides an explanation as to why the hexane-hexane CG RDFs were an underestimation compared to the atomistic reference RDFs, due to the potentials being less attractive, and why hexane-water interactions were overestimated in the CG RDFs, due to the potential being overly attractive.

Effect on increasing system size

In order to examine the effect of increasing atomistic system size on the performance of the IBI method on the hexane/water system, two larger systems have been studied. A second system (named HS2) contains 200 hexane molecules in water and a third system (named HS3) contains 400 hexane molecules in water. Both systems still have a hexane concentration of 11.6 wt.%, identical to HS1, but scaled up by a factor of 4 and 8. Contrary to HS1, visual inspection of both HS2 and HS3 CG simulation results shows the hexane phase separating to form a ball of hexane surrounded by water (figure 5.10, right). Qualitatively, the increase in system size has produced CG potentials better able to match the overall phase separation behaviour observed in the atomistic reference systems.

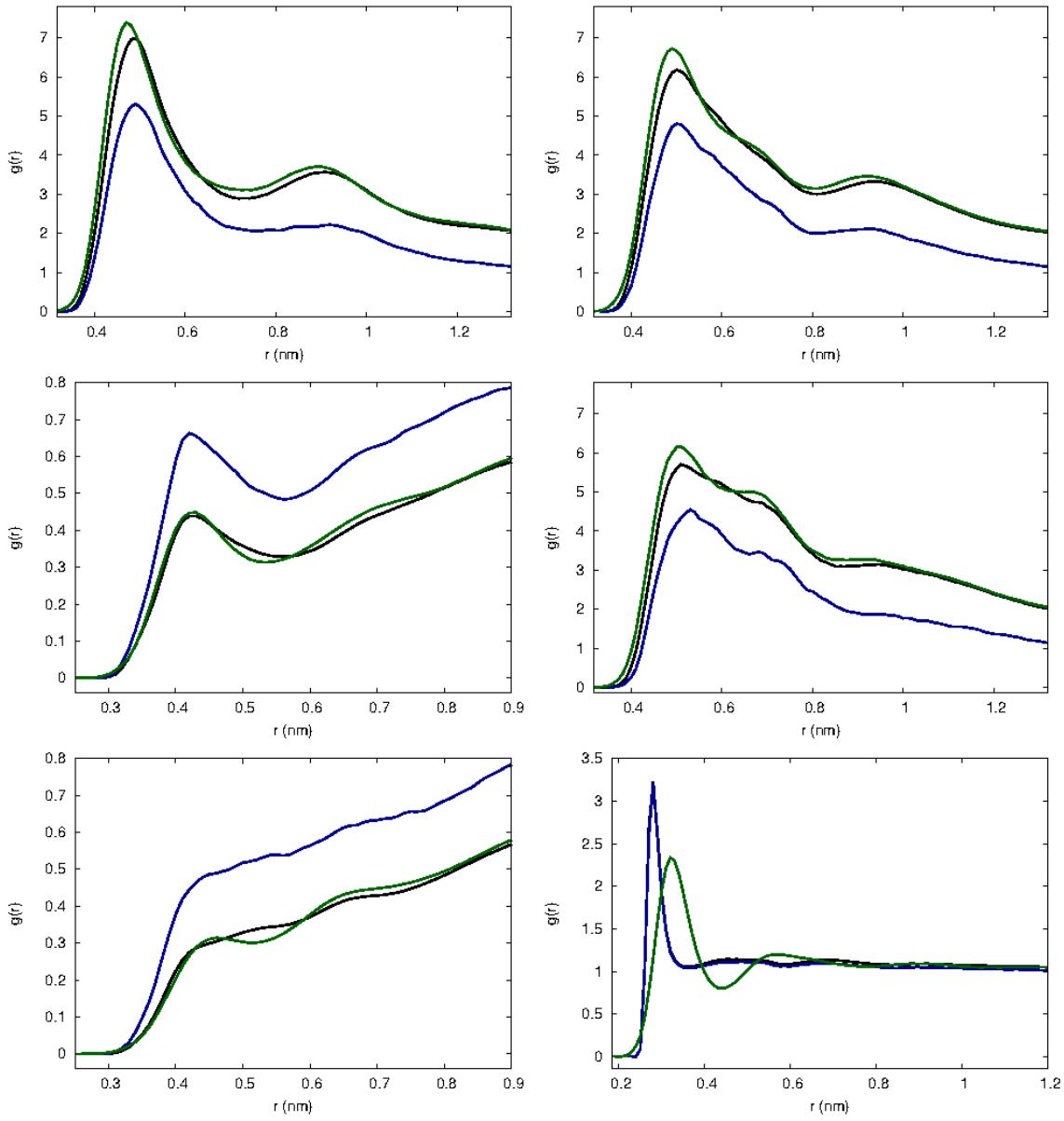

Figure 5.11: Radial distribution functions for the 6 non-bonded interactions within the coarse grained HS1 system: A-A (top left), A-B (top right), A-C (middle left), B-B (middle right), B-C (bottom left), C-C (bottom right), where atomistic (black), IBI (blue) and FM (green) distributions are shown.

An analysis of the RDF's for each interaction in HS2 can be seen in figure 5.12. Hexane-hexane interactions (A-A, A-B, B-B) shows an overestimation for the B-B interaction, but very good atomistic agreement for A-A and A-B interactions. The increased system size has generated more attractive CG potentials for these interactions compared to HS1, due to the surface effects of the hexane ball being decreased. The water-water potential (C-C) shows a perfect structural match on the primary peak, but fails to perfectly match at longer distances, with very small secondary and tertiary solvation shells being created by the CG potential. An increased system size has failed to perfectly reproduce the atomistic reference water-water interactions, despite these being structurally reproduced in HS1. Cross-term interactions (A-C, B-C) show an underestimation, in sharp contrast to the HS1 cross-term interactions.

An analysis of the RDF's for each interaction in HS3 can be seen in figure 5.13. Hexane-hexane interactions A-A and A-B show an underestimation, whilst BB has shown a structural profile not sufficiently close to the atomistic reference, with the primary peak being bimodal, and a secondary peak being found at long range. The water-water interaction (C-C) shows the worse match so far for water-water, with neither the primary peak nor the general shape being sufficiently reproduced. Cross-term interactions for hexane-water (A-C, B-C) are not only underestimates, but fail to reproduce the qualitative shape of the atomistic reference RDF at all. HS3 has shown to be quantitatively worse than HS2 at reproducing the structural features of the atomistic reference simulation.

IBI summary

IBI has been used to generate coarse grain potentials for an atomistic hexane/water mixture. Whilst homogeneous systems of pure hexane and pure water generated accurate CG potentials using the IBI method, a phase separated mixture of both components has proved to be much more complex to represent. For each system studied (HS1, HS2, HS3), IBI failed to fully converge all of the interactions to match the atomistic reference RDF's for that system.

A system size dependency is observed, with a sufficiently large size system size

being needed to fully describe the presence of both surface interactions of the hexane ball with water, and the hexane-hexane interactions in the centre of the ball. The systems showing the best results can be ordered as: HS2 > HS3 > HS1. While increasing the system size from HS1 proved to be effective, simply choosing the largest system was not the optimal solution, since HS3 proved to be worse to converge than HS2. Predicting a minimum size of system necessary would be extremely difficult, due to the need to then coarse grain a wide range of atomistic simulations of different sizes, and then compare the structural reproducibility. Essentially, this defeats the purpose of coarse graining a system to save computational time and work.

Hexane/water mixtures using FM

Coarse-grained potentials for HS1, HS2, and HS3 were also generated using the force matching method. Visual inspection of all 3 CG systems show a qualitative reproduction of the atomistic behaviour of hexane phase separating in water to form a ball. Comparison of the radial distribution functions for the CG system and the atomistic reference system for HS1 (figure 5.11) overall shows an extremely high degree of structural match between the two. Hexane-hexane (A-A, A-B, B-B) and hexane-water interactions (A-C, B-C) show almost perfect agreement in the RDF's, with only the water-water interaction (C-C) failing to sufficiently match the atomistic reference. The two larger system sizes of HS2 and HS3 both show near-identical structural behaviour for the CG results, with all hexane-hexane (A-A, A-B, B-B) and hexane-water interactions (A-C, B-C) all being lower in height than the atomistic reference. No improvement in the water-water interaction has been observed on moving from a smaller system size (HS1) to a larger system size (HS2 and HS3). It should be also noted that, in addition to these 3 systems, even in the pure water system, force matching failed to reproduce the identical structural characteristics of the atomistic simulation.

The effect of system size on the hexane/water mixtures shows an increase in system size not improving the structural accuracy from the CG potentials generated. Whilst all 3 systems qualitatively reproduce the phase separation behaviour, HS1

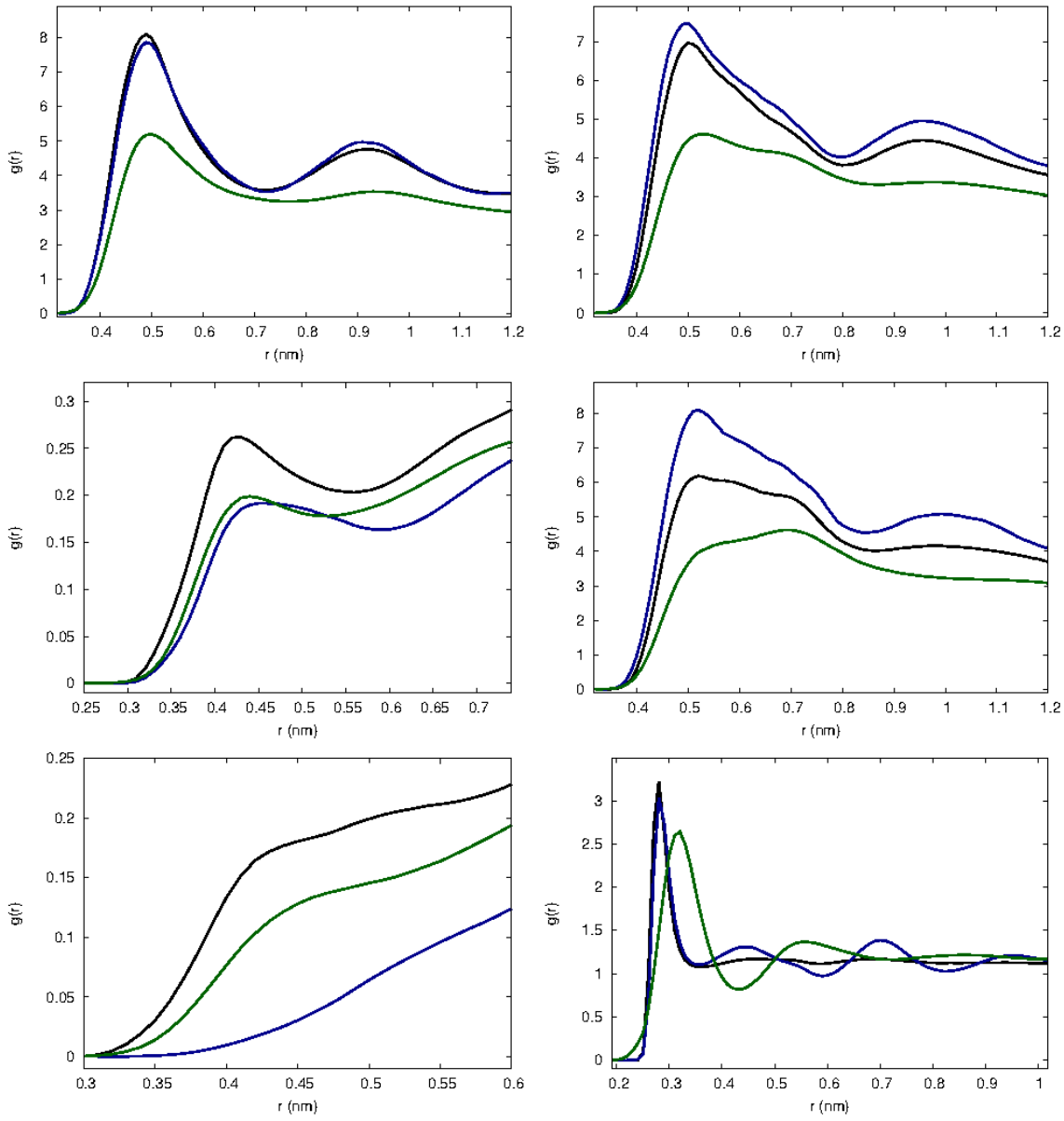

Figure 5.12: Radial distribution functions for the 6 non-bonded interactions within the coarse grained HS2 system: A-A (top left), A-B (top right), A-C (middle left), B-B (middle right), B-C (bottom left), C-C (bottom right), where atomistic (black), IBI (blue) and FM (green) distributions are shown.

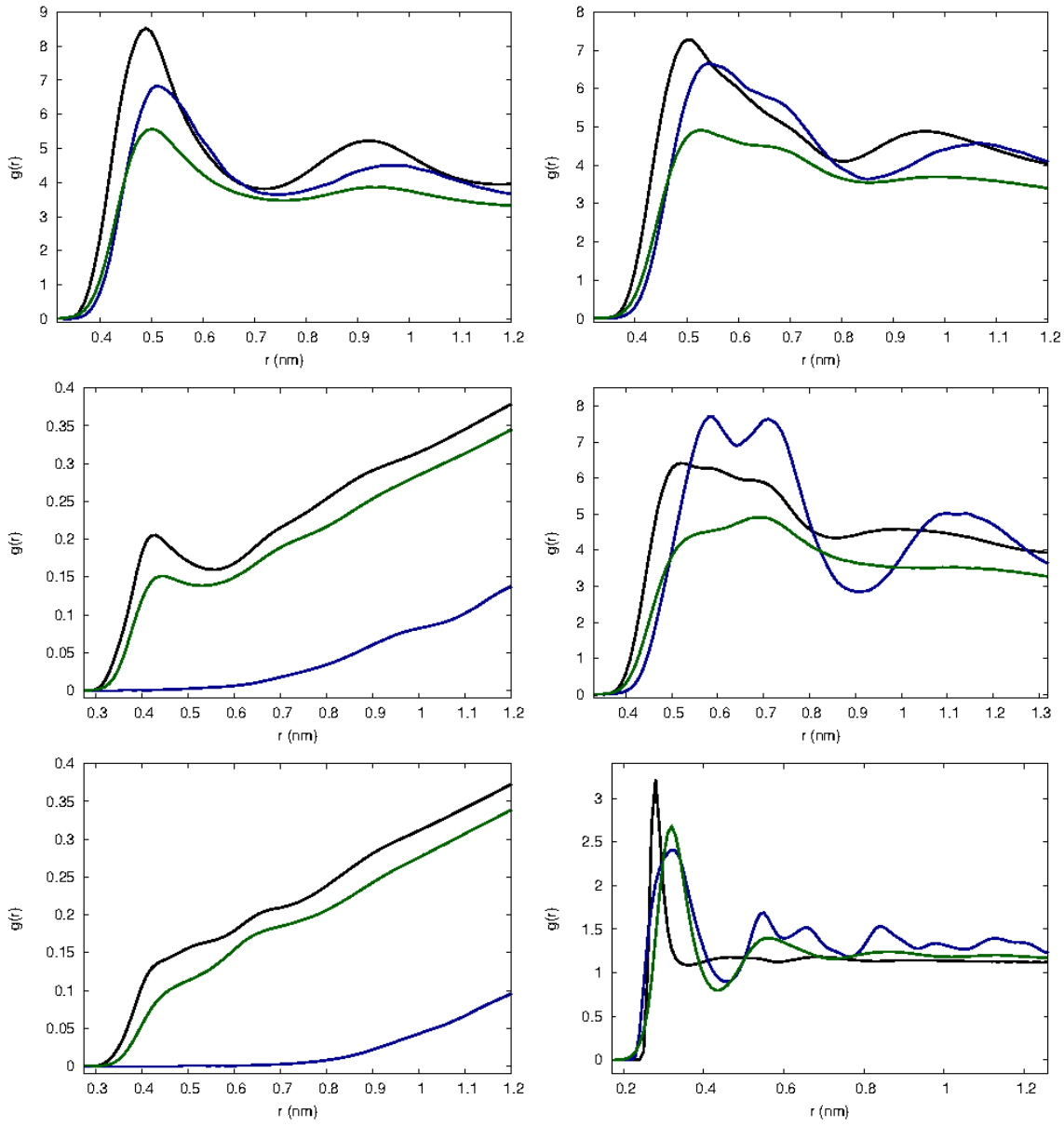

Figure 5.13: Radial distribution functions for the 6 non-bonded interactions within the coarse grained HS3 system: A-A (top left), A-B (top right), A-C (middle left), B-B (middle right), B-C (bottom left), C-C (bottom right), where atomistic (black), IBI (blue) and FM (green) distributions are shown.

shows the most accurate structural reproducibility compared to the atomistic reference system. HS2 and HS3 both show extremely similar structural characteristics, potentially meaning that there is a threshold by which increasing system size has no further effect (whether that be a negative or positive effect).

Conclusion

Coarse grain potentials for both the IBI and FM method on the hexane/water mixtures have been performed for three system sizes (HS1, HS2, HS3). Qualitatively, the FM method was the only method to produce the correct phase separation independent of system size, whilst IBI only achieved this at larger systems sizes (HS2 and HS3). A comparison of all the 6 CG systems studies shows the FM method for HS1 having the best structural performance.

The two CG methods can be compared by breaking down and examining the performance for 3 sets of interactions: hexane-hexane (A-A, A-B, B-B), hexane-water, (A-C, B-C), and water-water (C-C). Force matching for HS1 was the only system to generate correct potentials for all 3 of the hexane-hexane interactions. Whilst FM showed a good match for hexane-water interactions (HS1 and HS3), IBI never attained a good match in any of the 3 systems studied, showing a consistent inability to correctly deal with solute-solvent interactions in this phase separated system. On the other hand, whilst IBI has good matching for water-water interactions (HS1, HS2), FM fails to ever correctly reproduce the water-water radial distribution function for any of the 3 systems.

A comparison for the time required in generating potentials can also be made. The FM technique was substantially quicker than the IBI method. The IBI method requires simulation time for each iteration, which can be easily a minimum of a few hours per iteration when moving to larger systems such as HS3. Despite the large number of iteration steps needed for the IBI technique, it still failed to converge in this time.

In summary, whilst neither of the CG methods have been able to perfectly reproduce the atomistic reference system, many strengths and weaknesses of both the method and practical implementation of it, have been revealed by exploring these

hexane/water mixtures. Looking ahead to the coarse graining of a chromonic dye in solution, force matching appears to be a more promising method, given the more succesful results of dealing with a phase separated system.

5.4 Coarse graining of Dye A

Having explored the coarse graining of phase separated systems, with hexane/water mixtures, both the FM and IBI method can now be applied to a more complex phase separated system of an ionic chromonic dye in water. Mesogen Dye A, studied in Chapter 3, is an ideal candidate to coarse grain due to the wealth of atomistic simulations already completed and analysed. While atomistic simulations allowed short time scale aggregation behaviour to be observed (a few columns in solution), successfully coarse graining could potentially make possible the accurate study of many hundreds or thousands of aggregates in solution, allowing whole chromonic phases to be observed.

5.4.1 Structure and mapping

Two mapping schemes have been chosen for the Dye A chromonic system, a coarser system with many atoms grouped into single sites (labelled system "M1"), and a more refined system with less atoms per interaction site (labelled system "M2"). The structure and mapping schemes for Dye A can be seen in Fig 5.14.

A 5-site representation was used for M1 (Figure 5.14, top), whereby Dye A has 5 interaction sites: 2 symmetrical sites containing the outer methoxy and main arene core unit (named A), a single central site bridging these arene core units (named B), and 2 symmetrical sulphonate arms (named C). This representation provides 2 bond types (AB, BC), 3 angle types (ABC, ABA, CBC), no proper dihedrals, and 1 improper dihedral type (ABAC). The ethyl ammonium counterion has been reduced to 1 interaction site (D), and the solvent is a 1-site water (E). For this M1 representation, 15 non-bonded interactions are present in the system.

A 9-site representation was used for M2 (Figure 5.14, bottom), whereby Dye A has 9 interaction sites: 2 symmetrical outer sites for the methoxy groups (named A), 2 symmetrical end sulphonate groups (named B), 2 symmetrical alkyl chains of the sulphonate arms (named C), 2 symmetrical inner sites containing the arene core (named D), and a central site (named E). This representation provides 4 bond types (AD, DE, DC, BC), 5 angle types (ADE, DCB, ADC, DED, EDC), 4 proper dihedral

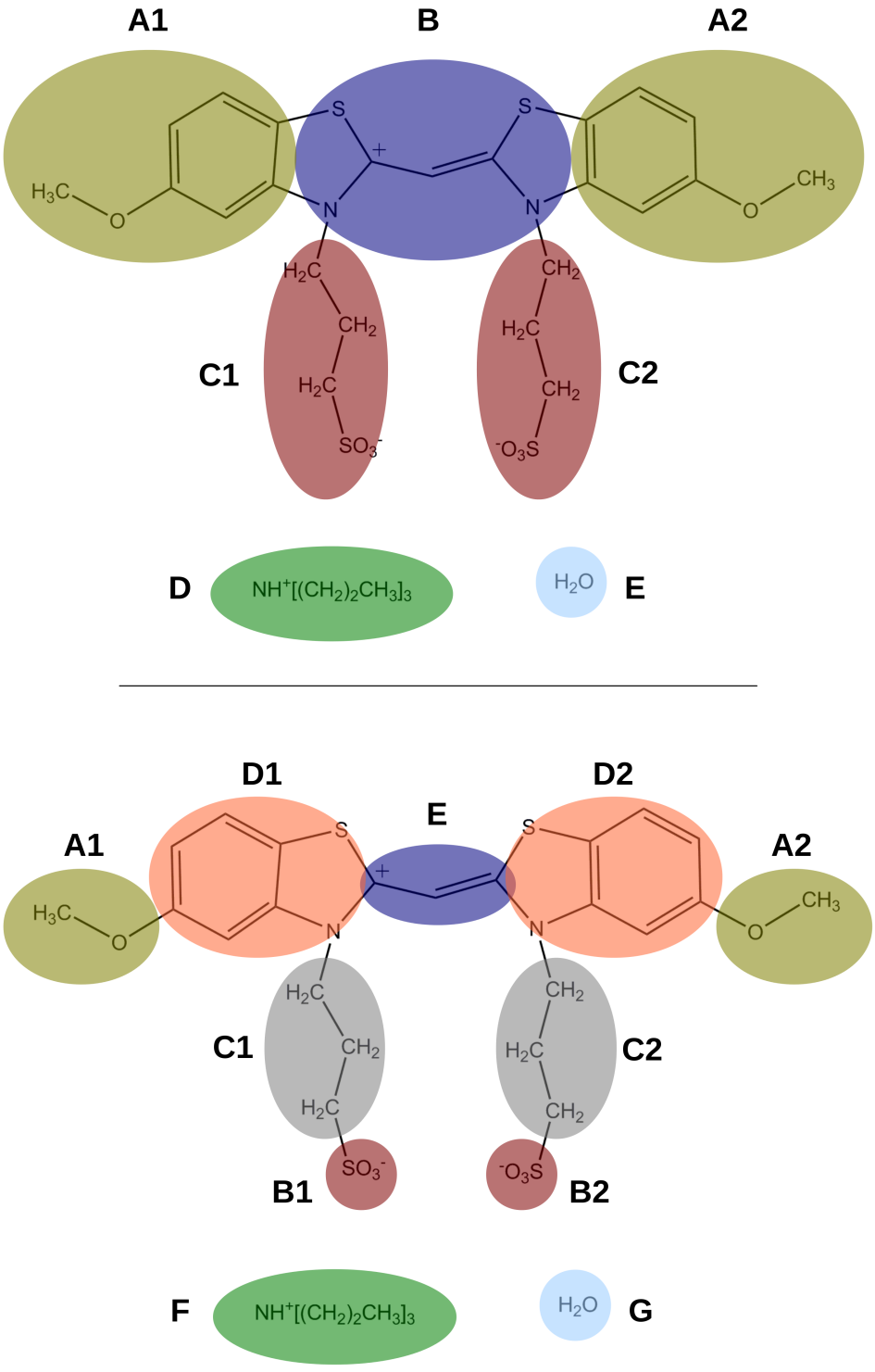

Figure 5.14: Two coarse-grained mapping schemes chosen for Dye A, with a coarser system labelled M1 (top) and a more refined system labelled M2 (bottom).

types (ADED, BCDA, BCDE, CDED), and 1 improper dihedral type (ADEC). The ethyl ammonium counterion has been reduced to 1 interaction site (F), and the solvent is a 1-site water (G). For this M2 representation, 28 non-bonded interactions are present in the system.

A hybrid coarse graining method will be applied, as with the previous hexane systems, whereby the bonded and non-bonded interactions are treated independently.

5.4.2 Bonded interactions

The bonded potentials have been generated by the Boltzmann inversion method applied to the atomistic reference systems. The bonded potentials for the M1 system can be seen in fig 5.15. The potential for bond AB shows a steeply rising potential with a minimum at 0.56 nm. The potential for bond BC is broader with a minimum at 0.62 nm, with the broadness attributable to the more dynamic motion of the sulphonate arms attached to the main core of the dye causing a wider distribution of bond distances. The angle ABA connects the rigid core of the dye molecule, with an angle of 180° showing a perfectly linear configuration between these interaction sites, while the potential for the angles connecting the rigid core to the sulphonate arms have minimum energy angles at 67-68° for ABC and 48° for CBC. An improper dihedral potential connecting interaction sites ABAC shows a global minimum at -95-105° and second minimum at 137°, with a barrier height of approximately 4.5 kJ/mol between these wells.

The bonded potentials for the M2 system can be seen in fig 5.16. The potentials for bonds AD and DE within the rigid core region show the minimum energy bond distances of 0.42 and 0.38 nm respectively. The potential for bond BC connecting the alkyl chain to the sulphonate head group shows an energy minimum at 0.3 nm. The potential for bond DC, corresponding to the linking bond between the core and flexible arm regions, has a broader energy well, with a minimum energy at 0.44 nm. The potentials for angles between interaction sites within the core, ADE and DED, show a strongly linear configuration with energy minimums at 154° and 143°. Interactions between core and flexible arms, such as DCB, ADC, EDC, have angle potentials with minima at 144°, 97° and 56-57° respectively. The most notable

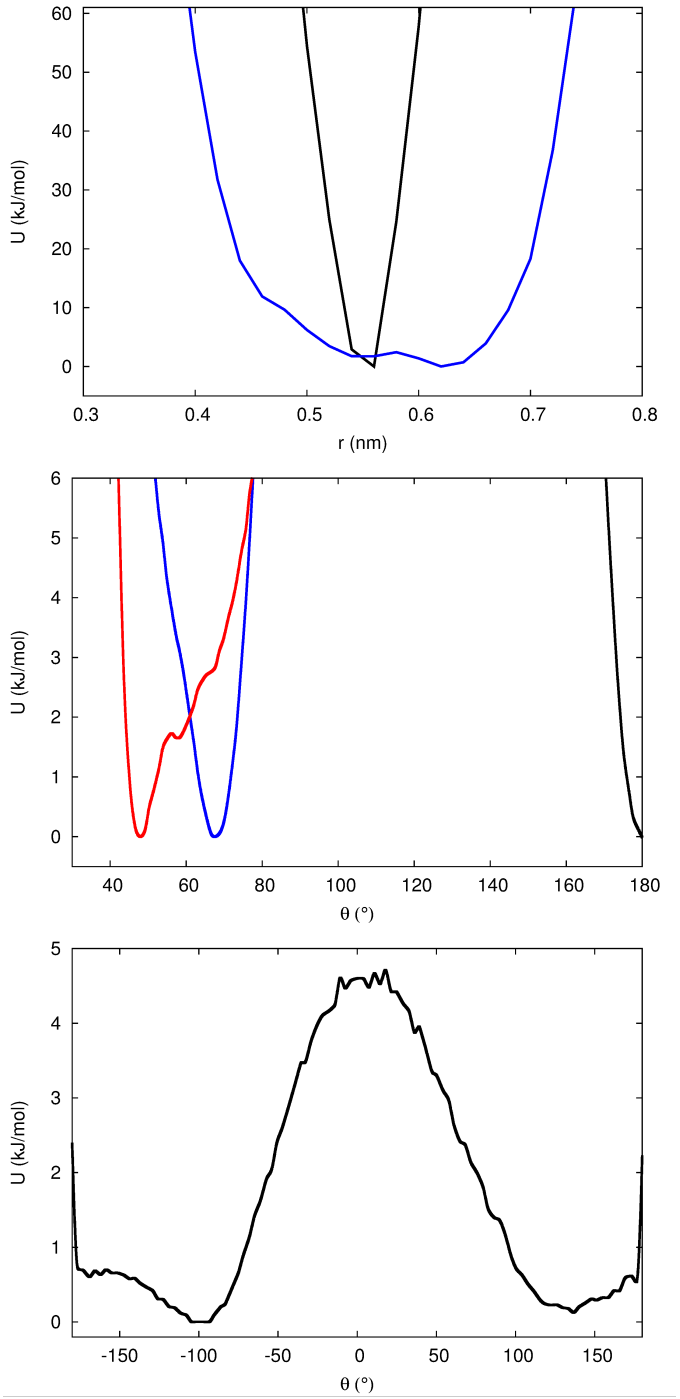

Figure 5.15: Bonded potentials for Dye A in the M1 system. This representation provides 2 bond types with AB (black) and BC (blue), 3 angle types (middle) with ABA (black), ABC (blue), CBC (red) and 1 improper dihedral type (bottom) ABAC (black).

dihedrals to mention are ADED and CDED, which impose rigidity across the central region of the dye molecule. Both dihedral potentials have a minimum at -176-177° and 176-177°, with a large barrier height of 25.5 kJ/mol between these wells. The flat region for both the ADED and CDED dihedral potentials is where the potential has been cut-off, due to this high energy region not ever been seen in the atomistic simulation.

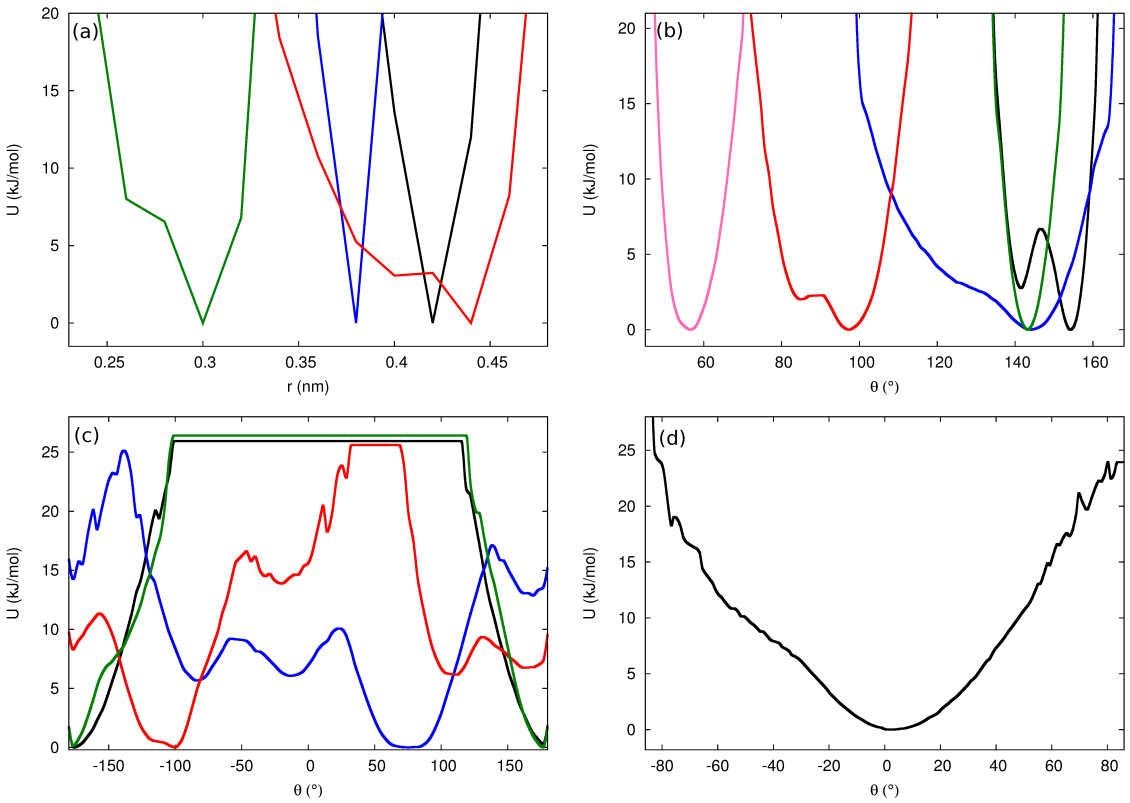

Figure 5.16: Bonded potentials for Dye A in the M2 system. This representation provides 4 bond types (a) with AD (black), DE (blue), DC (red), CB (green), 5 angle types (b) with ADE (black), DCB (blue), ADC (red), DED (green), EDC (pink), 4 dihedral types (c) with ADED (black), BCDA (blue), BCDE (red), CDED (green), and 1 improper dihedral type (d) with ABAC (black).

5.4.3 IBI of Dye A

The iterative Boltzmann inversion technique has been used to generate coarse-grained potentials for both the M1 and M2 system. Both simulation starting configurations is the equilibrated aggregate structure with the appropriate mapping scheme. While IBI failed to converge all potentials to the target radial distribution

functions, the iteration with the best overall convergence criteria match has been selected for both systems. After a minimum of 50 iterations, iteration 33 was selected for M1, and iteration 7 for M2. Further iteration steps from these selected iterations showed no more improvement in converging the potentials, with large fluctuations in convergence match criteria occuring for each iteration.

IBI of M1

The CG M1 system shows the chromonic aggregate retaining a stacked configuration for approximately 0.75 ns (figure 5.17, (a)), before permanently transitioning to a more clustered aggregate (figure 5.17, (b)). Whilst the reference atomistic simulation has the counterions freely dispersed in the water, the CG system shows the counterions being strongly attracted, and in effect binding, to the sulphonate arms (interaction site C) of the chromonic dye. Repulsion between the sulphonate arms between dye molecules (C-C interaction) is still retained, with interaction site C maintaining a maximum distance from each other in the aggregate. The CG bonded interactions work extremely well, with the core (A-B-A) retaining its planar structure seen in the atomistic simulations, and the core-sulphonate arms (B-C) moving more freely.

A structural comparison showing the radial distribution functions for both the atomistic simulation and the CG simulation can be seen for each interaction in figure 5.18. Overall, the IBI method failed to generate potentials that were capable of reconstructing the atomistic reference RDFs. However, several key interactions should be mentioned. Interactions between the core of the dye molecules (A-A, A-B, B-B) are moderately successful in reproducing the overall shape and the multiplicity of the RDFs, while other dye-dye interactions (A-C, B-C, C-C) were not sufficiently reproduced. All dye-counterion (A-D, B-D, C-D) and counterion-counterion (D-D) interactions have been reproduced extremely poorly. While the atomistic RDFs show the counterion dispersed in solution, the CG RDFs show a large degree of order and attraction, with even the overall shape of the reference RDF not being reproduced at all. Dye-water and water-water interactions have all converged relatively well, with the exception of D-E, where only the short range structure is managed well. In the

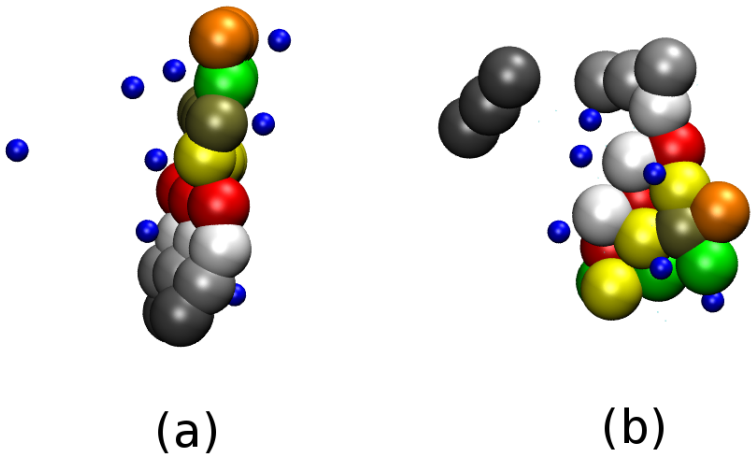

Figure 5.17: Snapshot from the IBI coarse grain simulation for M1, where (a) is the initial structure and (b) is the equilibrated structure. The cores of each dye molecule (beads A1, A2 and B) are shown in different colours for each dye molecule, and the counterion is blue. Water has been removed for visual clarity.

total 15 interactions in the system, the only RDF to be nearly perfectly reproduced was the water-water interaction (E-E). Here, it seems that converging 15 separate radial distribution functions is too challenging a task to accomplish within the IBI framework.

The clustering can be attributed to the coarseness of the mapping scheme of M1. The core unit interactions (A-A,B-B) and cross-terms (A-B) have long range order in the atomistic structure, and therefore CG potentials reflecting this order. Each dye molecule will be attracted to the neighbouring dye in the stack, thus explaining the initial retention of structure (a), but is also attracted to the next-neighbour in the column. Combining the effect of the counterions forming bridges between dye molecules via the sulphonate regions, allows structure (b) to form. The bridging can be demonstrated in (b), with the dark grey dye molecule having reoriented away from the central stack, but still connected to the aggregate via the counterion.

IBI of M2

The CG M2 system shows the dye molecules remaining in a stack for the entirety of the CG simulation. The central core of the dye molecules (D1-E-D2) form a linear

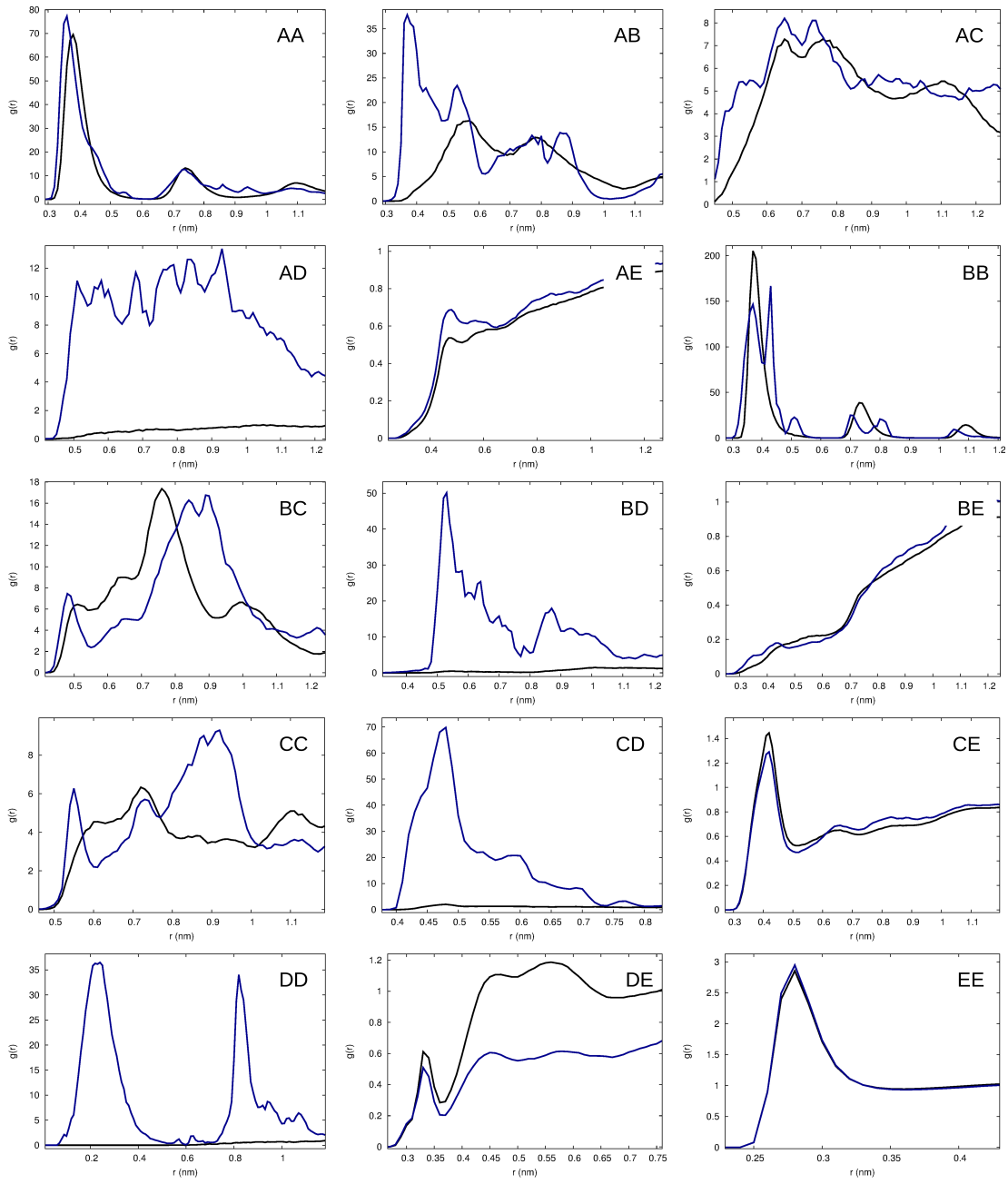

Figure 5.18: Radial distribution functions for CG M1 interactions from IBI, where black is the target distribution function and blue is the best match from the CG potential.

aggregate (fig 5.19, left), with the sulphonate arms (C-B) retaining an antiparallel configuration with respect to neighbours (fig 5.19, right). It should be noted that the core molecules are not perfectly overlapping, and appear to stagger slightly with respect to neighbours. Counterions are freely dispersed in solution, with no apparent affinity for binding to the dye molecule. The bonded interactions are very successful, with the core interaction sites (A1-D1-E-D2-A2) adopting a planar configuration from the atomistic reference system, and the sulphonate arms (interaction sites C-B) are more flexible with a wider range of motion.

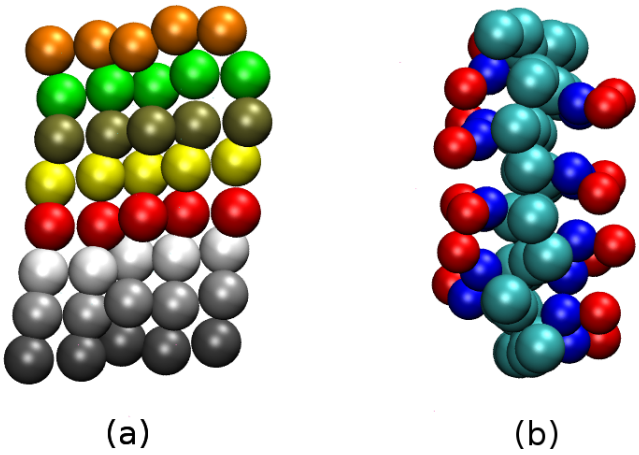

Figure 5.19: Snapshot from the IBI coarse grain simulation for M2. Left (a) shows the front view of the stack, with cores only shown and represented by different colours for each dye molecule. Right (b) shows the side profile of the stack, where the cores are highlighted in light blue, interaction sites C1 and C2 in dark blue, and B1 and B2 in red. The counterions and water molecules have been removed for visual clarity.

A structural comparison showing the radial distribution functions for both the atomistic simulation and the CG simulation can be seen for each interaction in figure 5.20. Overall, the IBI method failed to generate potentials that were capable of reconstructing the atomistic reference RDFs. However, several key interactions should be mentioned. Interactions between the core of the dye molecules appear to split into two groups, whereby interactions with a sharp, long-range order (A-A, D-D, E-E) are reproduced quite successfully, and the other RDFs (A-D, A-E, D-E) less so. The interactions between the sulphonate arms of the dye (B-B, C-C, B-C) are

generally poorly reproduced, with the exception of interaction B-B, which manages to crudely reproduce an overall similar shape to the target function. For the dye-arm interactions (A-B, A-C, B-D, C-D, B-E, C-E), the moderately successful interactions captured were A-B, C-D and B-D (this function is qualitatively good, but shifted to the right of the target function). All dye-counterion (A-F, B-F, C-F, D-F, E-F) and counterion-counterion (F-F) interactions have been reproduced quite poorly, whilst the effect of all these interactions still producing a system with the counterions dispersed in solution. From all the interactions involving water; Dye-water (A-G, B-G, C-G, D-G, E-G), counterion-water (F-G) and water-water interactions (G-G), the most successful interactions moderately reproduced were B-G, C-G, F-G, and G-G.

An analysis of the CG aggregate allows for a direct comparison to the atomistic reference aggregate. The stacking distance (distance between the cores of neighbouring dye molecules in the stack) is 0.37 nm for the CG aggregate, in comparison to the slightly shorter stacking distance of 0.35 nm for the atomistic system (figure 5.21, left). A twist angle analysis, determining the twisting angle between neighbouring dye molecules in a stack, shows that both systems adopt an antiparallel configuration with respect to the sulphonate arms (figure 5.21, right). The CG stack has a sharp peak at 177°, with a narrow distribution from 177-180°, whilst the atomistic has a broad peak from 162-168°, with a wide distribution of angles from 135-180°. This shows that the CG simulation has a more rigid overlap between neighbouring dye molecules, and allows for less of a twisting range. The long range helical properties of the aggregate can be analysed with a histogram showing population of twisting angle between all pairs, and distances along the stack (figure 5.22). Both systems show no long range twisting behaviour, with the CG simulation M2 showing more long range disorder than the atomistic reference system.

In summary, whilst the CG M2 simulation is qualitatively successful, the IBI method fails to structural reproduce each radial distribution function for each interaction. An analysis of the CG simulation aggregate shows a slightly larger stacking distance between dye molecules, as well as a narrower twisting angle range, and a more disordered long range behaviour through the stack.

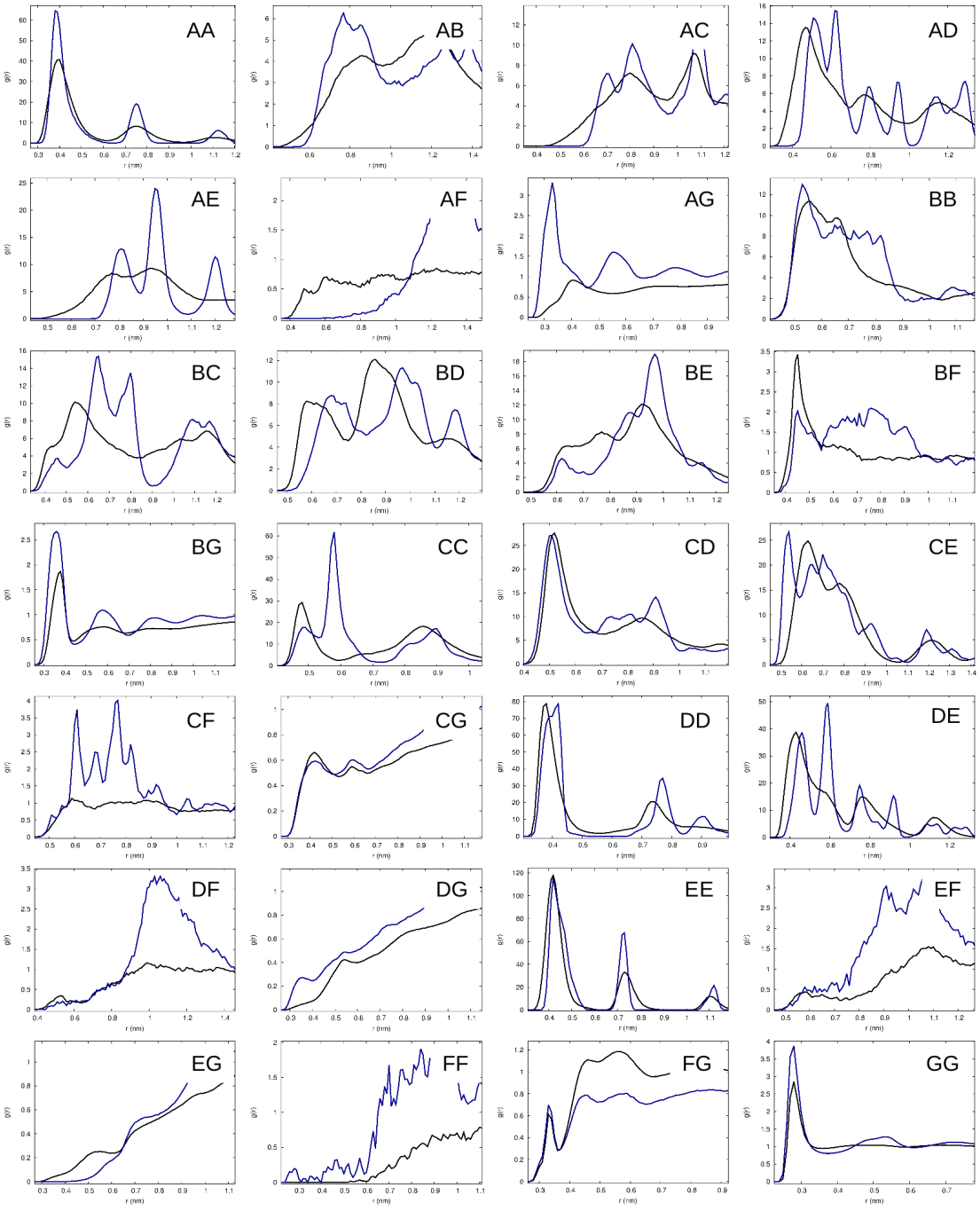

Figure 5.20: Radial distribution functions for CG M2 interactions, whereby black is the atomistic target distribution function and blue is the best match from the CG potential.

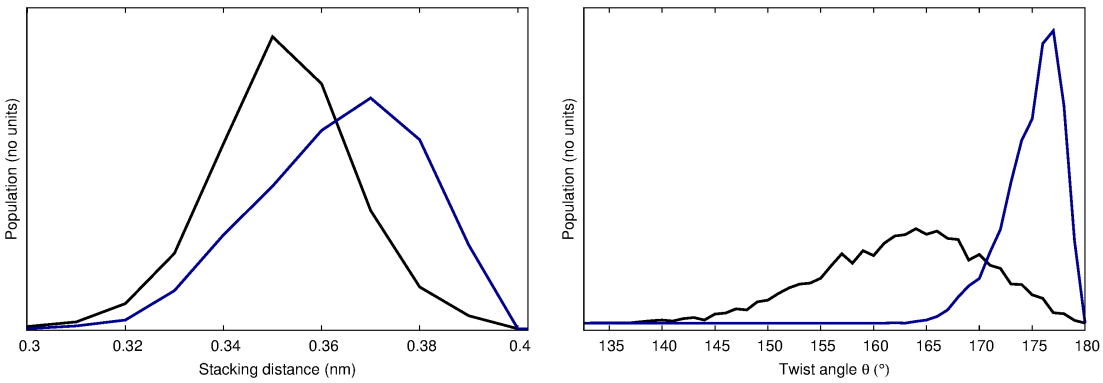

Figure 5.21: A comparison of the distance between neighbouring dye molecules (top) and the twist angle between neighbouring molecules (bottom) in a stack, with atomistic (black) and CG simulation M2 (blue).

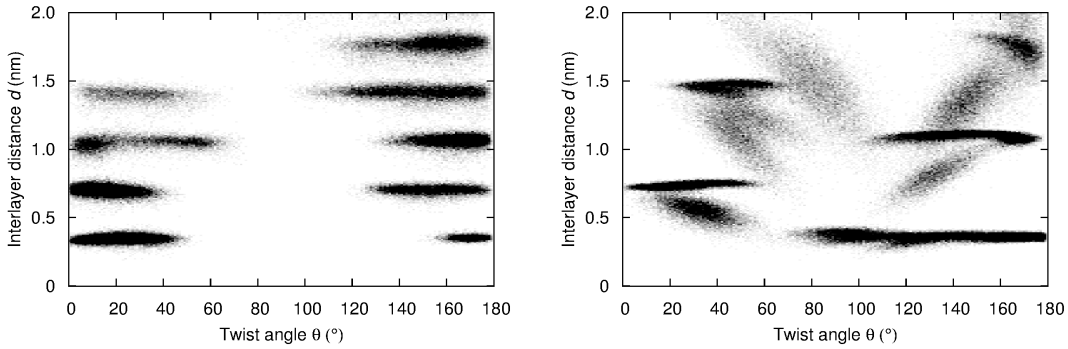

Figure 5.22: Histogram showing the population of twisting angle between all pairs and distances along the stack for the atomistic system (left) and the CG simulation M2 (right).

Improvements to the IBI method

The two main issues facing the successful application of the IBI method to chromonic Dye A is firstly, the number of non-bonded interactions in the CG system is simply too high to fit all successfully simultaneously, and secondly, the IBI method has been applied to an anisotropic system. The implementation of a potential solution to each of these problems is discussed now, with both being applied to mapping M2, the more successful of the two coarse graining schemes.

A possible approach to dealing with the large number of interactions that fail to converge, is to reduce the complexitiy of the target RDFs. This was achieved by restricting the target RDF to contain only the first peak (which is the most useful, short range information on the structure), and generating potentials to fit to this new, simplified RDF. A simplified target RDF approach proved unsuccessful, and did not increase the convergence of the CG RDFs to the new target RDFs.

An unconventional approach to deal with the second problem, whereby the IBI method is typically used for isotropic, not anistotropic systems, is to produce target RDFs from dispersed monomers and dimers in solution and not from an equilibrated stack. This was attempted by constructing an atomistic simulation identical to the atomistic reference previously used, but with the 8 dye molecules randomly dispersed in solution. The simulation was allowed to run for 5 ns, and then RDFs obtained from this time frame, in the hope of the RDFs losing the semi-crystalline nature seen in the equilibrated system. The non-equilibrated RDF approach proved unsuccessful, due to the dye molecules essentially binding on contact, and therefore producing semi-crystalline RDFs.

In the future it may be possible to use the initial coarse-grained model generated by IBI as a starting point for an improved coarse-grained model using group solubility data, as used in MARTINI, to refine the non-bonded potentials.

5.4.4 Force matching of Dye A

The force matching technique has been used in an attempt to generate suitable coarse grain potentials for M1 and M2. The force profiles, which show the force

as a function of distance between interaction sites, are generated and then converted into potentials. The force profiles for M1 can be seen in figure 5.23 and for M2 in figure 5.24. Both systems show a wide variety of force profiles, with key trends appearing for specific interactions in both systems. All dye-counterion and counterion-counterion interactions for M1 (A-D, B-D, C-D, D-D) and M2 (A-F, B-F, C-F, D-F, E-F, F-F) show extremely large forces with typical magnitudes being in the range of 10^7-10^{15} N. Counterion-water force profiles (D-E for M1, and F-G for M2) appear to be the only counterion interaction whereby the forces are not unreasonable. Several dye-dye interactions, such as B-B for M1, and E-E for M2, show forces that are significantly larger than expected. The similarity of the force profile issues occuring for both systems rule out the problem arising from the structural difference between M1 and M2, and can instead be attributed to a problem with the force matching technique itself on this type of system. The VOTCA force matching procedure appears to incorrectly deal with a combination of electrostatic and Van der Waals components of the total forces arising from the reference system. The presence of charged interaction sites in the system (for example, with the positively charged counterion) means that the Van der Waals potential could be in a very high energy region of the potential, but this is balanced by the coulombic interactions. However, a neglect of the coulombic component on the forces leads to these very unreasonable (and unrealistic) force profiles between many of the interactions in this ionic chromonic system.

Improvement methods for the force profiles

In an atomistic simulation, the total force experienced on a particular molecule or atom in the system will be a combination from both the electrostatic and Van der Waals components. The problem with VOTCA is that it only optimises the Van der Waals component. To overcome this problem and achieve sensible force profiles (and thus potentials) for the system, a method can be used that isolates the electrostatic component for the force. The electrostatic component was isolated by setting all σ and ϵ values in the atomistic topology file to zero, and then doing a force matching calculation on the same trajectory as the original. This results in forces that are

from the electrostatic component only. This force profile was then deducted from the from the original set generated, giving a force profile for the Van der Waals component only, as this is what VOTCA will optimise. Whilst theoretically this holds merit, the practical application of this method yielded no success with this ionic chromonic system. The resultant force profiles proved no more useful than the original set calculated.

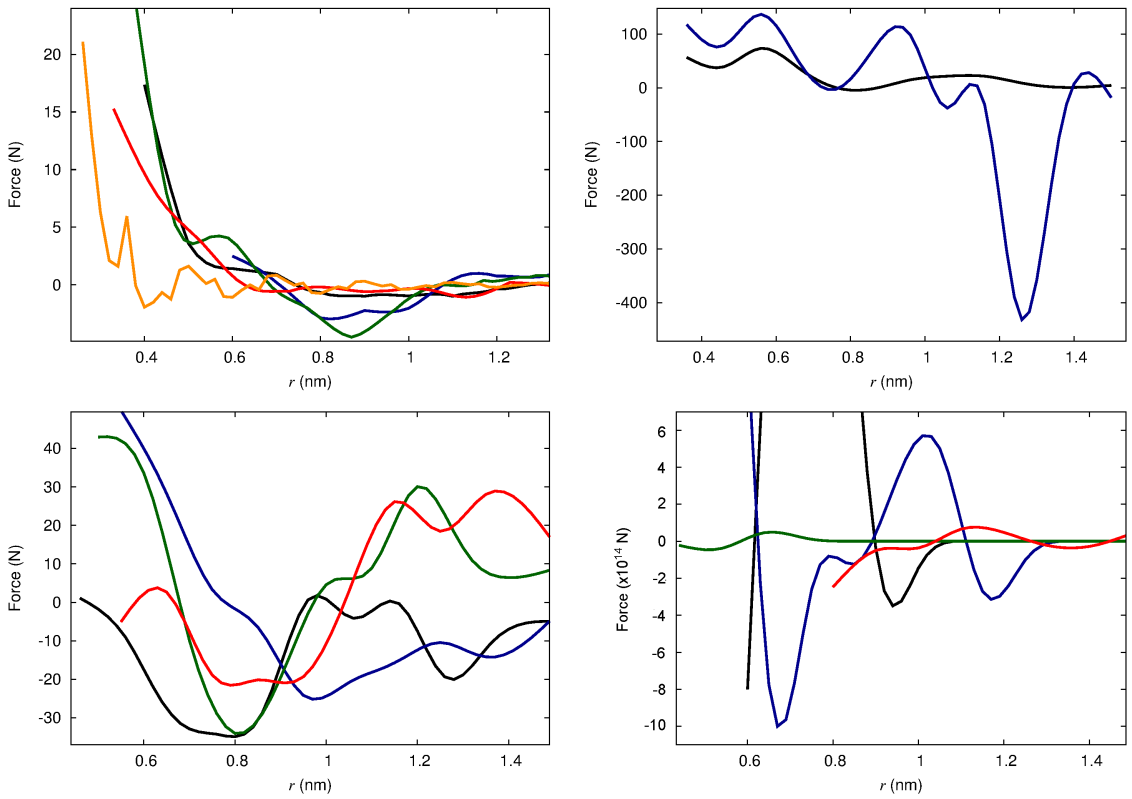

Figure 5.23: Force profiles for the non-bonded interactions within the M1 system. Separated into four plots for easier viewing, top left shows each interaction site with water: A-E (black), B-E (blue), C-E (green), D-E (red) and E-E (orange). Dye-dye interactions can be seen in top right with A-A (black) and B-B (blue), and bottom left with A-B (black), A-C (blue), B-C (green), C-C (red); and dye-counterion and counterion-counterion interactions in bottom right with A-D (black), B-D (blue), C-D (green) and D-D (red).

5.4.5 Conclusions

Systematic coarse graining techniques have been implemented in an attempt to coarse grain the chromonic Dye A in solution from a reference atomistic simulation. A hybrid scheme has been chosen, whereby the bonded and non-bonded interactions

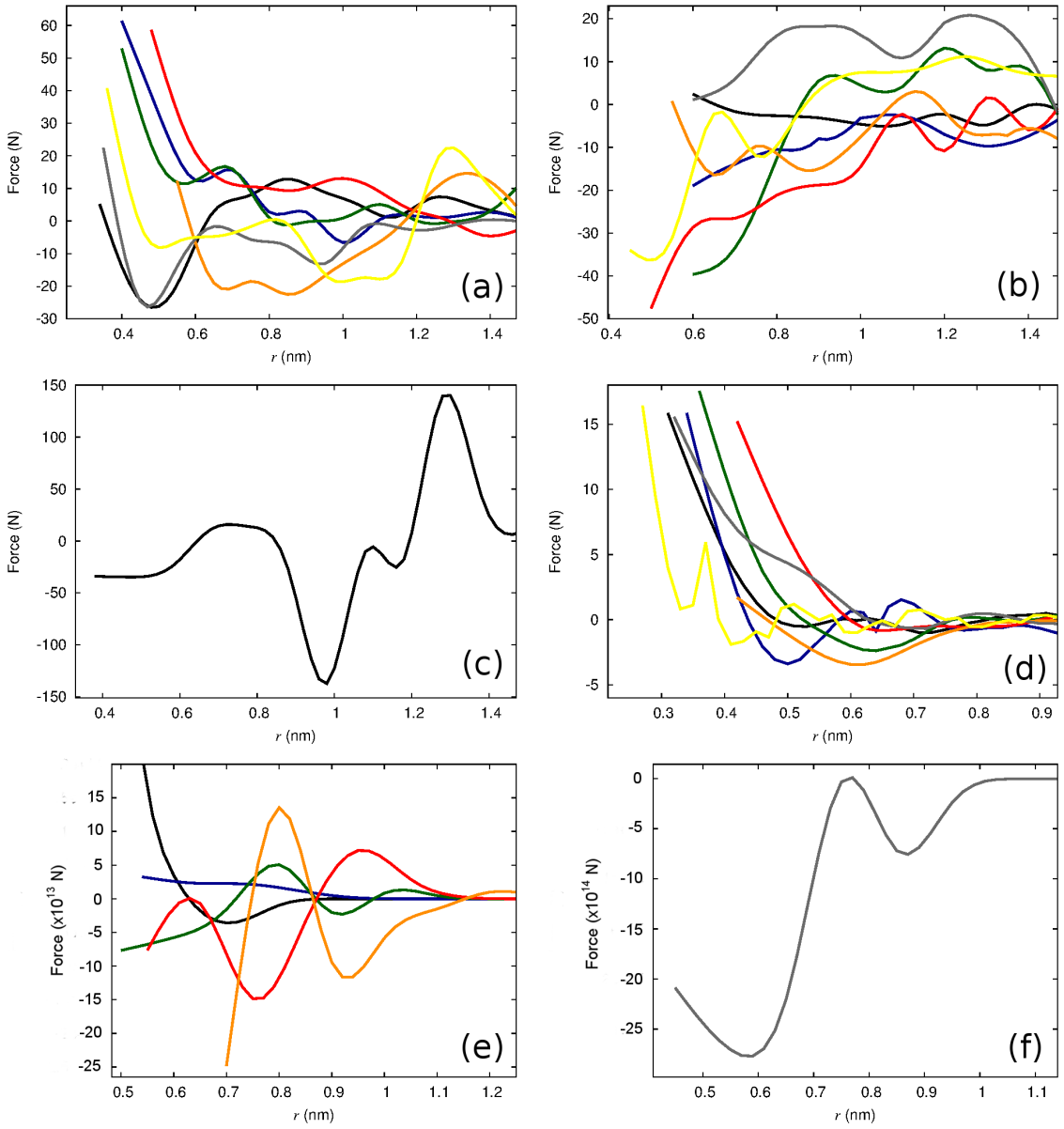

Figure 5.24: Force profiles for the non-bonded interactions within the M2 system. Separated into six plots for easier viewing, dye-dye interactions can be seen in (a) with A-A (black), A-D (blue), B-C (green), C-D (red), C-E (orange), D-D (grey), D-E (yellow), (b) with A-B (black), A-C (blue), A-E (green), B-B (red), B-D (orange), B-E (grey), C-C (yellow) and (c) with E-E (black). Interaction sites with water can be seen in (d) with A-G (black), B-G (blue), C-G (green), D-G (red), E-G (orange), F-G (grey), G-G (yellow). Dye-counterion and counterion-counterion interactions can be seen in (e) with B-F (black), C-F (blue), D-F (green), E-F (red), F-F (orange), and (f) with A-F (grey).

are treated independently. The bonded interactions have been generated by a simple Boltzmann inverstion (BI), and the non-bonded potentials have been generated by either iterative Boltzmann inversion (IBI) or force-matching (FM). Two mapping schemes were selected for Dye A: a 5-site model (labelled M1) and a 9-site model (labelled M2).

The IBI method showed markedly different results based on the coarseness of the mapping scheme used. The more refined coarse grain mapping scheme (M2) was the only system to yield a stacked aggregate for the dye molecules, qualitatively matching the atomistic (despite not quantitatively reproducing the structure accurately). The M1 system proved to have too coarse a mapping scheme, with the dye molecules and associated counterions aggregating into a cluster. The mapping scheme had an obvious effect not only on the aggregation behaviour of the dye molecules, but also the counterion behaviour in solution. Despite M2 showing a moderately successful result, the IBI method failed to generate potentials to match the target RDFs due to the number of interactions in the system being too large to successfully fit. Several alternative methods to overcome this issue, namely using simplified RDF functions (cut-off function after first shell) and taking structural statistics from a dispersed system of monomers and dimers, proved ineffective at obtaining a better convergence.

The FM method, as implemented in the VOTCA package, neglected to properly account for the coulombic components of the atomistic forces when screening interactions, and therefore provided unrealistical force profiles for both M1 and M2 systems, with extremely large force magnitudes for many of the interactions in the system. Manually isolating the coulombic component and removing this from the original force profile proved unsuccessful in generating potentials of any reasonable magnitude, let alone potentials reproducing the atomistic reference system.

In summary, both *bottom-up* methods attemped, the structure-based approach (IBI) and the hybrid force-matching approach (FM), proved unsuccessful in generating suitable potentials capable of adequately capturing the behaviour of the reference atomistic system of chromonic mesogen Dye A in solution.

A *Top-down* approach where experimental data is used to guide the parame-

terisation of the coarse-grained potentials, such as the MARTINI force field[123] or SAFT,[124] could prove more effective at reproducing the behaviour of chromonic dyes in solution and successfully capturing the aggregation behaviour.

Chapter 6

Summary

The work presented in this thesis concerns the simulation of ionic chromonic liquid crystals. We have shown that molecular simulations can be a powerful additional tool to study chromonic liquid crystals, complementing experimental results and theoretical predictions. Atomistic modelling, in particular, can aid in forming a molecular level understanding of the chromonic aggregation process.

Chapter 3: Atomistic simulations were used to study the aggregation behaviour of chromonic Dye A and Dye B. Experimental literature for Dye A suggested a double-width column structure for the nematic, while simulations showed a single-width chiral column forming from achiral mesogens. Higher concentrations were proposed to have a layer structure for the smectic, with a repeat in the intralayer structure every 5 rows. Molecular simulation indeed observed a layer structure, but with columns aligning and adopting an achiral structure, with no interdigitation. Key experimental results (such as interlayer distance, intralayer distance etc) could be reproduced, and in some cases reinterpreted based on the molecular level picture the simulation provided. Experimental literature for Dye B also suggested a layer structure. Simulation results showed low concentrations of Dye B forming single-width columns with no chirality, whilst higher concentrations showed a layer structure. Considering the high degree of structural overlap between Dye A and Dye B, this study also demonstrates how sensitive the aggregation process is to minor changes in the molecular structure of the mesogen.

Chapter 4: Atomistic simulations were used to study the aggregation behaviour of the chromonic molecule known as the Bordeaux dye. Experimental literature for the Bordeaux dye suggested a double-width column, however, simulation results showed the presence of a single-width column only. An extensive thermodynamic analysis of the free energy changes upon aggregation showed a strongly enthalpy-driven process, with a binding energy far greater than that calculated in the experimental literature. The free energy landscape showed the presence of two energy minima, with the global minimum showing a packing mode of a H-aggregate, and the local minimum a J-aggregate. This suggests that the Bordeaux dye could have a more complex model of aggregation than a simple H aggregate, with Y-junctions and shifted junctions causing the larger than expected cross sectional area. Contrary to other chromonic systems studied in this thesis, a high degree of order occurs between the counterion and the sulphonate region of the dye molecule. It has not been investigated to what extent this effects the binding between molecules, or whether the counterions can act as a bridge between neighbouring columns in solution, however, this is an interesting topic that requires more examination.

Chapter 5: A *bottom-up* coarse graining approach was used to study phase separated mixtures. The force matching (FM) and iterative Boltzmann inversion (IBI) techniques were initally applied to hexane/water mixtures of three different system sizes. IBI showed an obvious system size dependency, with qualitative reproduction of the atomistic system only achieved at larger system sizes. FM qualitatively showed no system size dependency, but structural analysis showed only the smallest system size reproducing the reference system accurately. Both the FM and IBI techniques were then applied on the more complex chromonic Dye A system, whereby two mapping schemes were selected and studied. The FM technique showed complications dealing with an ionic system, with a secondary charge-only FM profile needing to be deducted from the original FM profiles. This still resulted in unrealistic force profiles. Within the IBI framework, it seems that converging such a large number of separate radial distribution functions is too challenging. Whilst several problems and issues have been highlighted, and attempted to be addressed,

a successful coarse-grained model for chromonic Dye A in solution was not achieved with either technique.

This thesis presents an extensive computational study of ionic chromonics in solution. A range of atomistic and coarse graining techniques have been employed in an attempt to capture different properties that occur on different time and length scales.

Atomistic simulations have proven to be particularly effective at capturing the short range behaviour of chromonic aggregation. Simulation results for Dye A, Dye B and the Bordeaux dye not only show the typical face-to-face aggregation of dye molecules to form columns, but is also able to reproduce key experimental findings, such as structural features (X-ray distances, cross sectional areas) and thermodynamic properties (binding energies). The formation of a chiral column for Dye A and an achiral column for Dye B at low concentrations demonstrates how atomistic modelling is highly sensitive to even small changes in the molecular structure. The layer structure at high concentrations for Dye A and Dye B is particularly exciting, due to these simulations providing molecular level detail of stacking arrangements within the layer, which has been a complicated structure for experimentalists to rationalise. The Bordeaux dye exhibits potential counterion bridges in the system, showing that in addition to the molecular structure of the dye, the counterion could potentially have a large effect on the aggregation behaviour. Future work could entail using atomistic simulation to probe the stacking behaviour of an extensive library of chromonic mesogens, with a range of counterions, to strongly refine our knowledge of the structure-property relationship.

Bottom-up coarse graining techniques for ionic chromonics, with Dye A as the selected target molecule, proved much less fruitful in generating successful results in comparison to atomistic simulations. IBI is extremely difficult to scale, because as the number of different sites increase, the number of RDFs to fit at the same time becomes prohibitively expensive. For *bottom-up* techniques, improving on the current force matching method appears to be the most promising avenue to explore. One thing that could be done is to modify the VOTCA force matching program to

work with basis sets of analytical potentials, instead of numerical potentials, which seem to have been numerically unstable in the fitting procedure for the ionic Dye A system.

Recent studies has shown that the best realistic approach for designing coarse graining models of chromonic systems is using *top-down* coarse graining techniques, such as SAFT and MARTINI. The SAFT method, as well as the MARTINI method, have been employed on chromonic systems[135] and aromatic molecules,[136–138] and appears to be a promising approach, particularly when specific molecular interactions need to be captured. A possible method to expand MARTINI specifically for a chromonic system could be by fitting free transfer free energies of different groups.

A big area of future study needs to be going beyond simulating single chromonic stacks, to simulating chromonic phases, allowing to connect with experimental studies of bulk nematic and other phases. In theory, good coarse grained models could achieve this, but currently the only progress has been at the DPD level.[139,140] The ability to accurately simulate chromonic phase behaviour could deeply enrich our understanding of chromonics, and could answer questions such as how the mysterious P and O phase is structured,[1] and why some chromonics form at incredibly low concentrations.

The use of simulation to better understand chromonic systems is on the rise, and is already showing great potential in furthering our understanding of chromonic behaviour.

Bibliography

[1] J Lydon. Chromonic review. *Journal of Material Chemistry*, 20(45):10071–10099, 2010.

[2] GNI Clark, CD Cappa, JD Smith, RJ Saykally, and T Head-Gordon. The structure of ambient water. *Molecular Physics*, 108(11):1415–1433, 2010.

[3] CJ Bottrill. *Dye-Dye interactions in aqueous mesophases*. PhD thesis, University of Manchester, 1999.

[4] CJ Bottrill, WJ Harrison, DL Mateer, and GJT Tiddy. Chromonic liquid crystals formed by the cyanine dyes 5,5'-dimethoxy-bis-(3,3'-di-sulphopropyl)-thiacyanine triethylammonium salt and 5,5'-dichloro-bis-(3,3'-di-sulphopropyl)-thiacyanine triethylammonium salt in water, studied by optical microscopy, x-ray diffraction and nmr. *to be published*.

[5] MR Tomasik and PJ Collings. Aggregation behavior and chromonic liquid crystal phase of a dye derived from naphthalenecarboxylic acid. *Journal of Physical Chemistry B*, 112(32):9883–9889, 2008.

[6] HS Park, SW Kang, L Tortora, Y Nastishin, D Finotello, S Kumar, and OD Lavrentovich. Self-assembly of lyotropic chromonic liquid crystal sunset yellow and effects of ionic additives. *Journal of Physical Chemistry B*, 112(51):16307–16319, 2008.

[7] LM Blinov. *Structure and properties of liquid crystals*, volume 123. Springer Science & Business Media, 2010.

[8] S Chandrasekhar. Liquid crystals. *Cambridge University, Cambridge*, 1992.

[9] F Reinitzer. Beiträge zur kenntniss des cholesterins. *Monatshefte für Chemie*, 9(1):421–441, 1888.

[10] F Reinitzer. Contributions to the knowledge of cholesterol. *Liquid Crystals*, 5(1):7–18, 1989.

[11] TJ Sluckin, DA Dunmur, and H Stegemeyer. *Crystals that flow: Classic papers from the history of liquid crystals*. Taylor & Francis London, 2004.

[12] H Kelker. History of liquid crystals. *Molecular Crystals and Liquid Crystals*, 21(1-2):1–48, 1973.

[13] H Kelker. Survey of the early history of liquid crystals. *Molecular Crystals and Liquid Crystals*, 165(1):1–43, 1988.

[14] JW Goodby, D Demus, J Goodby, GW Gray, H-W Spiess, and V Vill. Symmetry and chirality in liquid crystals. *Handbook of Liquid Crystals: Fundamentals, Volume 1*, pages 115–132, 1998.

[15] E Priestly. *Introduction to liquid crystals*. Springer Science & Business Media, 2012.

[16] S Chandrasekhar and GS Ranganath. Discotic liquid crystals. *Reports on Progress in Physics*, 53(1):57, 1990.

[17] TK Attwood, JE Lydon, C Hall, and GJT Tiddy. The distinction between chromonic and amphiphilic lyotropic mesophases. *Liquid Crystals*, 7(5):657–668, 1990.

[18] GJT Tiddy. Surfactant-water liquid crystal phases. *Physics reports*, 57(1):1–46, 1980.

[19] AMF Neto and SRA Salinas. *The Physics of Lyotropic Liquid Crystals: Phase Transitions and Structural Properties: Phase Transitions and Structural Properties*. OUP Oxford, 2005.

[20] K Holmberg, DO Shah, and MJ Schwuger. *Handbook of applied surface and colloid chemistry*, volume 1. Wiley New York, 2002.

[21] J Lydon. Chromonic mesophases. *Current Opinion in Colloid and Interface Science*, 8(6):480–490, 2004.

[22] TL Crowley, C Bottrill, D Mateer, WJ Harrison, and GJT Tiddy. Lyotropic chromonic liquid crystals: neutron scattering studies of shear-induced orientation and reorientation. *Colloids and Surfaces A*, 129:95–115, 1997.

[23] WJ Harrison, DL Mateer, and GJT Tiddy. J-aggregates and liquid crystal structures of cyanine dyes. *Faraday Discussions*, 104:139–154, 1996.

[24] TK Attwood, JE Lydon, and F Jones. The chromonic phases of dyes. *Liquid Crystals*, 1(6):499–507, 1986.

[25] GJT Tiddy, DL Mateer, AP Ormerod, WJ Harrison, and DJ Edwards. Highly ordered aggregates in dilute dye-water systems. *Langmuir*, 11(2):390–393, 1995.

[26] L Onsager. The effects of shape on the interaction of colloidal particles. *Annals of the New York Academy of Sciences*, 51(1):627–659, 1949.

[27] MA Bates and D Frenkel. Phase behavior of two-dimensional hard rod fluids. *The Journal of Chemical Physics*, 112(22):10034–10041, 2000.

[28] R Zwanzig. First-order phase transitions in a gas of long thin rods. *The Journal of Chemical Physics*, 39(7):1714–1721, 1963.

[29] H Sandquist. Anisotropic aqueous solution. *Berichte der Bunsen-Gesellschaft Physical Chemistry Chemical Physics*, 48:2054–2055, 1915.

[30] H Freundlich, R Stern, and H Zocher. The colloidal chemistry of arsphenamine and neoarsphenamine. *Biochemische Zeitschrift*, 138:307–317, 1923.

[31] SW Tam-Chang and L Huang. Chromonic liquid crystals: properties and applications as functional materials. *Chemical Communications*, 17:1957–1967, 2008.

[32] T Fujiwara and K Ichimura. Surface-assisted photoalignment control of lyotropic liquid crystals. part 2. Photopatterning of aqueous solutions of a water-soluble anti-asthmatic drug as lyotropic liquid crystals. *Journal of Materials Chemistry*, 12(12):3387–3391, 2002.

[33] K Ichimura. *Molecular Cyrstals and Liquid Crystals*, 298:221, 1997.

[34] D Matsunaga, T Tamaki, H Akiyama, and K Ichimura. Photofabrication of micro-patterned polarizing elements for stereoscopic displays. *Advanced Materials*, 14(20):1477–1480, 2002.

[35] SV Shiyanovskii, OD Lavrentovich, T Schneider, T Ishikawa, II Smalyukh, CJ Woolverton, GD Niehaus, and KJ Doane. Lyotropic chromonic liquid crystals for biological sensing applications. *Molecular Crystals and Liquid Crystals*, 434(1):259–587, 2005.

[36] CJ Woolverton, E Gustely, L Li, and OD Lavrentovich. Liquid crystal effects on bacterial viability. *Liquid Crystals*, 32(4):417–423, 2005.

[37] SV Shiyanovskii, T Schneider, II Smalyukh, T Ishikawa, GD Niehaus, KJ Doane, CJ Woolverton, and OD Lavrentovich. Real-time microbe detection based on director distortions around growing immune complexes in lyotropic chromonic liquid crystals. *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 71(2):020702, 2005.

[38] MS Dresselhaus and IL Thomas. Alternative energy technologies. *Nature*, 414(6861):332–337, 2001.

[39] A Tolkki, E Vuorimaa, V Chukharev, H Lemmetyinen, P Ihalainen, J Peltonen, V Dehm, and F Würthner. Langmuir-Schaeffer films from a pi-pi stacking perylenediimide dye: organization and charge transfer properties. *Langmuir*, 26(9):6630–6637, 2009.

[40] SW Tam-Chang, J Helbley, TD Carson, W Seo, and IK Iverson. Template-guided organization of chromonic liquid crystals into micropatterned anisotropic organic solids. *Chemical Communications*, 5:503–505, 2006.

[41] AR Leach. *Molecular modelling: principles and applications*. Pearson education, 2001.

[42] MP Allen and DJ Tildesley. *Computer simulation of liquids*. Oxford university press, 1989.

[43] VR Horowitz, LA Janowitz, AL Modic, PA Heiney, and PJ Collings. Aggregation behavior and chromonic liquid crystal properties of an anionic monoazo dye. *Physical Review E*, 72(4):041710, 2005.

[44] L Joshi, SW Kang, DM Agra-Kooijman, and S Kumar. Concentration, temperature, and pH dependence of sunset-yellow aggregates in aqueous solutions: An x-ray investigation. *Physical Review E*, 80(4):041703, 2009.

[45] SK Prasad, GG Nair, G Hegde, and V Jayalakshmi. Evidence of wormlike micellar behavior in chromonic liquid crystals: Rheological, x-ray, and dielectric studies. *Journal of Physical Chemistry B*, 111(33):9741–9746, 2007.

[46] DJ Edwards, JW Jones, O Lozman, AP Ormerod, M Sintyureva, and GJT Tiddy. Chromonic liquid crystal formation by Edicol Sunset Yellow. *Journal of Physical Chemistry B*, 112(46):14628–14636, 2008.

[47] F Chami and MR Wilson. Molecular order in a chromonic liquid crystal: a molecular simulation study of the anionic azo dye sunset yellow. *Journal of the American Chemical Society*, 132(22):7794–7802, 2010.

[48] N Boden, RJ Bushby, L Ferris, C Hardy, and F Sixl. Designing new lyotropic amphiphilic mesogens to optimize the stability of nematic phases. *Liquid Crystals*, 1(2):109–125, 1986.

[49] N Boden, RJ Bushby, C Hardy, and F Sixl. Phase behaviour and structure of a non-ionic discoidal amphiphile in water. *Chemical Physics Letters*, 123(5):359–364, 1986.

[50] LM Ferris. PhD thesis, University of Leeds, 1989.

[51] SD Towlson. PhD thesis, University of Leeds, 1990.

[52] JF Hubbard. PhD thesis, University of Leeds, 1997.

[53] A Akinshina, M Walker, MR Wilson, GJT Tiddy, AJ Masters, and P Carbone. Thermodynamics of the self-assembly of non-ionic chromonic molecules using

atomistic simulations. the case of TP6EO2M in aqueous solution. *Soft Matter*, 11(4):680–691, 2015.

[54] D Frenkel and B Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Academic press, 2001.

[55] T Darden, D York, and L Pedersen. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[56] IG Tironi, R Sperb, PE Smith, and WF van Gunsteren. A generalized reaction field method for molecular dynamics simulations. *Journal of Chemical Physics*, 102(13):5451–5459, 1995.

[57] HQ Ding, N Karasawa, and WA Goddard III. Atomic level simulations on a million particles: The cell multipole method for Coulomb and London nonbond interactions. *Journal of Chemical Physics*, 97(6):4309–4315, 1992.

[58] MJ Robertson, J Tirado-Rives, and WL Jorgensen. Improved peptide and protein torsional energetics with the opls-aa force field. *Journal of Chemical Theory and Computation*, 11(7):3499–3509, 2015.

[59] C Oostenbrink, A Villa, AE Mark, and WF Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The gromos force-field parameter sets 53a5 and 53a6. *Journal of Computational Chemistry*, 25(13):1096–987X, 2004.

[60] J Wang, RM Wolf, JW Caldwell, PA Kollman, and DA Case. Development and testing of a general Amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.

[61] L Verlet. Computer" experiments" on classical fluids. i. Thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159(1):98, 1967.

[62] RW Hockney, SP Goel, and JW Eastwood. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics*, 14(2):148–158, 1974.

[63] WC Swope, HC Andersen, PH Berens, and KR Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *Journal of Chemical Physics*, 76(1):637–649, 1982.

[64] HJC Berendsen. Transport properties computed by linear response through weak coupling to a bath. In *Computer Simulation in Materials Science*, pages 139–155. Springer, 1991.

[65] DJ Evans and BL Holian. The Nosé–Hoover thermostat. *Journal of Chemical Physics*, 83(8):4069–4074, 1985.

[66] S Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[67] WG Hoover. Canonical dynamics: equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985.

[68] HJC Berendsen, JPM van Postma, WF van Gunsteren, ARHJ DiNola, and JR Haak. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics*, 81(8):3684–3690, 1984.

[69] M Parrinello and A Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981.

[70] JM Kincaid and JJ Weis. Radial distribution function of a hard-sphere solid. *Molecular Physics*, 34(4):931–938, 1977.

[71] E Waisman. The radial distribution function for a fluid of hard spheres at high densities: mean spherical integral equation approach. *Molecular Physics*, 25(1):45–48, 1973.

[72] JG Kirkwood and EM Boggs. The radial distribution function in liquids. *Journal of Chemical Physics*, 10(6):394–402, 1942.

[73] JL Yarnell, MJ Katz, Ro Go Wenzel, and SH Koenig. Structure factor and radial distribution function for liquid argon at 85 K. *Physical Review A*, 7(6):2130, 1973.

[74] A Villa, C Peter, and NFA van der Vegt. Self-assembling dipeptides: conformational sampling in solvent-free coarse-grained simulation. *Physical Chemistry Chemical Physics*, 11(12):2077–2086, 2009.

[75] A Villa, NFA van der Vegt, and C Peter. Self-assembling dipeptides: including solvent degrees of freedom in a coarse-grained model. *Physical Chemistry Chemical Physics*, 11(12):2068–2076, 2009.

[76] C Li, J Shen, C Peter, and NFA van der Vegt. A chemically accurate implicit-solvent coarse-grained model for polystyrenesulfonate solutions. *Macromolecules*, 45(5):2551–2561, 2012.

[77] RM Neumann. Entropic approach to brownian movement. *American Journal of Physics*, 48(5):354–357, 1980.

[78] WL DeLano. PyMOL, 2002.

[79] WL DeLano. The PyMOL molecular graphics system. *http://pymol. org*, 2002.

[80] DA Case, TE Cheatham, T Darden, H Gohlke, R Luo, KM Merz, A Onufriev, C Simmerling, B Wang, and RJ Woods. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16):1668–1688, 2005.

[81] S Pronk, S Páll, R Schulz, P Larsson, P Bjelkmar, R Apostolov, MR Shirts, JC. Smith, PM. Kasson, D van der Spoel, B Hess, and E Lindahl. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, 2013.

[82] B Hess, C Kutzner, D Van Der Spoel, and E Lindahl. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.

[83] D Van Der Spoel, E Lindahl, B Hess, G Groenhof, AE Mark, and HJC Berendsen. GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, 26(16):1701–1718, 2005.

[84] E Lindahl, B Hess, and D Van Der Spoel. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, 7(8):306–317, 2001.

[85] HJC Berendsen, D van der Spoel, and R van Drunen. GROMACS: a message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, 1995.

[86] MJ Frisch, GW Trucks, Hs B Schlegel, GE Scuseria, MA Robb, JR Cheeseman, JA Montgomery Jr, TKKN Vreven, KN Kudin, JC Burant, et al. Gaussian 03, revision C. 02; Gaussian, Inc. *Wallingford, CT*, 26, 2004.

[87] AD Becke. Becke's three parameter hybrid method using the LYP correlation functional. *Journal of Chemical Physics*, 98:5648–5652, 1993.

[88] C Lee, W Yang, and RG Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37:785–789, Jan 1988.

[89] B Hess, H Bekker, HJC Berendsen, JGEM Fraaije, et al. LINCS: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[90] P Mark and L Nilsson. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *Journal of Physical Chemistry A*, 105(43):9954–9960, 2001.

[91] B Jójárt and TA Martinek. Performance of the general Amber force field in modelling aqueous POPC membrane bilayers. *Journal of Computational Chemistry*, 28(12):2051–2058, 2007.

[92] AJ Dickinson, ND LaRacuente, CB McKitterick, and PJ Collings. Aggregate structure and free energy changes in chromonic liquid crystals. *Molecular Crystals and Liquid Crystals*, 509(1):9–751, 2009.

[93] CB McKitterick, NL Erb-Satullo, ND LaRacuente, AJ Dickinson, and PJ Collings. Aggregation properties of the chromonic liquid crystal benzopurpurin 4B. *The Journal of Physical Chemistry B*, 114(5):1888–1896, 2010.

[94] HV Berlepsch, K Ludwig, and C Böttcher. Pinacyanol chloride forms mesoscopic H-and J-aggregates in aqueous solution–a spectroscopic and cryo-transmission electron microscopy study. *Physical Chemistry Chemical Physics*, 16(22):10659–10668, 2014.

[95] H Von Berlepsch, C Böttcher, and L Dähne. Structure of J-aggregates of pseudoisocyanine dye in aqueous solution. *Journal of Physical Chemistry B*, 104(37):8792–8799, 2000.

[96] H Von Berlepsch, C Böttcher, A Ouart, C Burger, S Dähne, and S Kirstein. Supramolecular structures of J-aggregates of carbocyanine dyes in solution. *Journal of Physical Chemistry B*, 104(22):5255–5262, 2000.

[97] C Didraga, A Pugžlys, PR Hania, H von Berlepsch, K Duppen, and J Knoester. Structure, spectroscopy, and microscopic model of tubular carbocyanine dye aggregates. *Journal of Physical Chemistry B*, 108(39):14976–14985, 2004.

[98] H von Berlepsch and C Böttcher. Network superstructure of pseudoisocyanine J-aggregates in aqueous sodium chloride solution revealed by cryo-transmission electron microscopy. *Journal of Physical Chemistry B*, 106(12):3146–3150, 2002.

[99] J Parkash, JH Robblee, J Agnew, E Gibbs, P Collings, RF Pasternack, and JC de Paula. Depolarized resonance light scattering by porphyrin and chlorophyll a aggregates. *Biophysical journal*, 74(4):2089–2099, 1998.

[100] A Speranza and P Sollich. Simplified onsager theory for isotropic–nematic phase equilibria of length polydisperse hard rods. *Journal of Chemical Physics*, 117(11):5421–5436, 2002.

[101] PJ Camp and MP Allen. Hard ellipsoid rod-plate mixtures: Onsager theory and computer simulations. *Physica A: Statistical Mechanics and its Applications*, 229(3-4):410–427, 1996.

[102] X Qiu, K Andresen, LW Kwok, JS Lamb, HY Park, and L Pollack. Inter-DNA attraction mediated by divalent counterions. *Physical review letters*, 99(3):038104, 2007.

[103] VA Bloomfield. DNA condensation. *Current Opinion in Structural Biology*, 6(3):334–341, 1996.

[104] E Raspaud, D Durand, and F Livolant. Interhelical spacing in liquid crystalline spermine and spermidine-DNA precipitates. *Biophysical journal*, 88(1):392–403, 2005.

[105] Y Burak, G Ariel, and D Andelman. Competition between condensation of monovalent and multivalent ions in DNA aggregation. *Current Opinion in Colloid and Interface Science*, 9(1):53–58, 2004.

[106] L Dai, Y Mu, L Nordenskiöld, and JRC van der Maarel. Molecular dynamics simulation of multivalent-ion mediated attraction between DNA molecules. *Physical review letters*, 100(11):118301, 2008.

[107] J Faraudo and A Travesset. The many origins of charge inversion in electrolyte solutions: effects of discrete interfacial charges. *Journal of Physical Chemistry C*, 111(2):987–994, 2007.

[108] DK Chattoraj, LC Gosule, and JA Schellman. DNA condensation with polyamines: Ii. Electron microscopic studies. *Journal of Molecular Biology*, 121(3):327–337, 1978.

[109] J Pelta, D Durand, J Doucet, and F Livolant. DNA mesophases induced by spermidine: structural properties and biological implications. *Biophysical Journal*, 71(1):48–63, 1996.

[110] AY Grosberg, TT Nguyen, and BI Shklovskii. Colloquium: the physics of charge inversion in chemical and biological systems. *Reviews of Modern Physics*, 74(2):329, 2002.

[111] A Naji, S Jungblut, AG Moreira, and RR Netz. Electrostatic interactions in strongly coupled soft matter. *Physica A: Statistical Mechanics and its Applications*, 352(1):131–170, 2005.

[112] I Borukhov, RF Bruinsma, WM Gelbart, and AJ Liu. Structural polymorphism of the cytoskeleton: a model of linker-assisted filament aggregation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(10):3673–3678, 2005.

[113] M Walker, AJ Masters, and MR Wilson. Self-assembly and mesophase formation in a non-ionic chromonic liquid crystal system: insights from dissipative particle dynamics simulations. *Physical Chemistry Chemical Physics*, 16(42):23074–23081, 2014.

[114] PB Warren. Dissipative particle dynamics. *Current Opinion in Colloid and Interface Science*, 3(6):620–624, 1998.

[115] V Tozzini. Coarse-grained models for proteins. *Current Opinion in Structural Biology*, 15(2):144–150, 2005.

[116] A Koliński. Protein modeling and structure prediction with a reduced representation. *Acta Biochimica Polonica*, 51, 2004.

[117] AYL Sim, P Minary, and M Levitt. Modeling nucleic acids. *Current Opinion in Structural Biology*, 22(3):273–278, 2012.

[118] S Takada. Coarse-grained molecular simulations of large biomolecules. *Current Opinion in Structural Biology*, 22(2):130–137, 2012.

[119] MG Martin and JI Siepmann. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *Journal of Physical Chemistry B*, 102(14):2569–2577, 1998.

[120] L Yang, CH Tan, MJ Hsieh, J Wang, Y Duan, P Cieplak, J Caldwell, PA Kollman, and R Luo. New-generation Amber united-atom force field. *Journal of Physical Chemistry B*, 110(26):13166–13176, 2006.

[121] C Oostenbrink, A Villa, AE Mark, and WF Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry*, 25(13):1656–1676, 2004.

[122] JF Dama, AV Sinitskiy, M McCullagh, J Weare, B Roux, AR Dinner, and GA Voth. The theory of ultra-coarse-graining. 1. General principles. *Journal of Chemical Theory and Computation*, 9(5):2466–2480, 2013.

[123] SJ Marrink, HJ Risselada, S Yefimov, DP Tieleman, and AH De Vries. The MARTINI force field: coarse grained model for biomolecular simulations. *Journal of Physical Chemistry B*, 111(27):7812–7824, 2007.

[124] WG Chapman, KE Gubbins, G Jackson, and M Radosz. New reference equation of state for associating liquids. *Industrial and Engineering Chemistry Research*, 29(8):1709–1721, 1990.

[125] CA López, AJ Rzepiela, AH De Vries, L Dijkhuizen, PH Hünenberger, and SJ Marrink. Martini coarse-grained force field: extension to carbohydrates. *Journal of Chemical Theory and Computation*, 5(12):3195–3210, 2009.

[126] L Monticelli, SK Kandasamy, X Periole, RG Larson, DP Tieleman, and SJ Marrink. The MARTINI coarse-grained force field: extension to proteins. *Journal of Chemical Theory and Computation*, 4(5):819–834, 2008.

[127] A Gautieri, A Russo, S Vesentini, A Redaelli, and MJ Buehler. Coarse-grained model of collagen molecules using an extended MARTINI force field. *Journal of Chemical Theory and Computation*, 6(4):1210–1218, 2010.

[128] W Tschöp, K Kremer, J Batoulis, T Bürger, and O Hahn. Simulation of polymer melts. i. Coarse-graining procedure for polycarbonates. *Acta Polymerica*, 49(2-3):61–74, 1998.

[129] V Rühle, C Junghans, A Lukyanov, K Kremer, and D Andrienko. Versatile object-oriented toolkit for coarse-graining applications. *Journal of Chemical Theory and Computation*, 5(12):3211–3223, 2009.

[130] D Reith, M Pütz, and F Müller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *Journal of computational chemistry*, 24(13):1624–1636, 2003.

[131] S Izvekov and GA Voth. A multiscale coarse-graining method for biomolecular systems. *Journal of Physical Chemistry B*, 109(7):2469–2473, 2005.

[132] F Ercolessi and JB Adams. Interatomic potentials from first-principles calculations: the force-matching method. *Europhysics Letters*, 26(8):583, 1994.

[133] S Izvekov, M Parrinello, CJ Burnham, and GA Voth. Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching. *Journal of Chemical Physics*, 120(23):10896–10913, 2004.

[134] V Rühle and C Junghans. Hybrid approaches to coarse-graining using the VOTCA package: Liquid hexane. *Macromolecular Theory and Simulations*, 20(7):472–477, 2011.

[135] TD Potter, J Tasche, EL Barrett, M Walker, and MR Wilson. Development of new coarse-grained models for chromonic liquid crystals: insights from top-down approaches. *Liquid Crystals*, 44(12-13):1979–1989, 2017.

[136] TF Headen, ES Boek, G Jackson, TS Totton, and EA Müller. Simulation of asphaltene aggregation through molecular dynamics: Insights and limitations. *Energy & Fuels*, 31(2):1108–1125, 2017.

[137] JF Jover, EA Müller, AJ Haslam, A Galindo, G Jackson, H Toulhoat, and C Nieto-Draghi. Aspects of asphaltene aggregation obtained from coarse-grained molecular modeling. *Energy & Fuels*, 29(2):556–566, 2015.

[138] T Lafitte, C Avendaño, V Papaioannou, A Galindo, CS Adjiman, G Jackson, and EA Müller. Saft-γ force field for the simulation of molecular fluids: 3. coarse-grained models of benzene and hetero-group models of n-decylbenzene. *Molecular Physics*, 110(11-12):1189–1203, 2012.

[139] M Walker and MR Wilson. Formation of complex self-assembled aggregates in non-ionic chromonics: dimer and trimer columns, layer structures and spontaneous chirality. *Soft Matter*, 12(41):8588–8594, 2016.

[140] M Walker, AJ Masters, and MR Wilson. Self-assembly and mesophase formation in a non-ionic chromonic liquid crystal system: insights from dissipative particle dynamics simulations. *Physical Chemistry Chemical Physics*, 16(42):23074–23081, 2014.